THE GENERATION OF REGIONAL PRONUNCIATIONS OF ENGLISH FOR SPEECH SYNTHESIS[1]

Susan Fitt
e-mail sue@cstr.ed.ac.uk
Centre for Speech Technology Research
University of Edinburgh
80 South Bridge
Edinburgh
UK

ABSTRACT

Most speech synthesisers and recognisers for English currently use pronunciation lexicons in standard British or American accents, but as use of speech technology grows there will be more demand for the incorporation of regional accents. This paper describes the use of rules to transform existing lexicons of standard British and American pronunciations to a set of regional British and American accents. The paper briefly discusses some features describes of the regional accents in the project, and the framework used for generating pronunciations. Certain theoretical and practical problems are highlighted; for some of these, solutions are suggested, but it is shown that some difficulties cannot be resolved by automatic rules. However, although the method described cannot produce phonetic transcriptions with 100% accuracy, it is more accurate than using letter-to-sound rules, and faster than producing transcriptions by hand.

1. INTRODUCTION

For some applications of speech synthesis, and for some users, output in standard accents is inappropriate, and as the use of speech systems increases there will be an increase in demand for regional accents of English. Access to regional pronunciation variants will also be of value for speech recognition systems. A labour-efficient way of producing these is needed; this paper describes the production by rule of pronunciation lexicons for five accents of English, using as input the information already contained in a lexicon of standard British and American pronunciations. There is the added benefit that since many linguistic rules are used by more than one accent, the ground-work is laid for producing further accents.

2. REGIONAL ACCENTS

Three British accents were chosen (as spoken in Edinburgh, Cardiff and Leeds, to represent Scottish, Welsh and Northern English), and two American ones (New York and South Carolina, to represent Eastern and Southern American); regional features were based primarily on the descriptions in [1], with native-speaker input where possible. The regional accents are abbreviated in this paper as: Br(Sc) = Edinburgh; Br(W) = Cardiff; Br(N) = Leeds; Am(E) = New York; and Am(S) = South Carolina. For the standard accents, Br(RP) = RP, and Am(Gen) = General American.

The accents generated represent fairly educated regional speech, though some optional rules were included which produce broader accents. The division between 'obligatory' and 'optional' rules is somewhat artificial, as there may be speakers from the region who have a noticeably local accent but do not use all of the 'obligatory' rules as their speech is somewhat closer to the standard accent. However, it enables us to produce pronunciation lexicons which represent the main features of the regional accents, while allowing some freedom of variation.

Some examples of the regional characteristics to be included in a lexicon, i.e. excluding such features as rhythm and intonation, are given below. (Throughout this paper, transcriptions are given in IPA unless otherwise specified.)

Feature	**Example**	**Br(RP)**	**Br(Sc)**
Rhoticity	'horse'	/hɔs/	/hɔɹs/
Vowel length/ quality distinctions	'tide' 'tied'	/taɪd/ /taɪd/	/tʌid/ /taˑed/

Figure 1: Some features of Edinburgh English

Feature	**Example**	**Br(RP)**	**Br(W)**
Presence of /ɬ/	'llewelyn'	/luˈɛ.lɪn/	/ɬuˈɛ.lɪn/
Full vowel in final syllables	'endless'	/ˈɛnd.ləs/	/ˈɛnd.lɛs/

Figure 2: Some features of Cardiff English

[1]This work was supported by France Telecom CNET under the contract 6RC0328.

Feature	Example	Br(RP)	Br(N)
Different use of /æ/ - /ɑ/ (realised in Leeds as /a/ - /aː/)	'hat' 'dance' 'part'	/hæt/ /dɑns/ /pɑt/	/hat/ /dans/ /paːt/
Optional /h/-dropping	'hot'	/hɒt/	/ɒt/

Figure 3: Some features of Leeds English

Feature	Example	Am(Gen)	Am(E)
Presence of /ɪu/	'new'	/nu/	/nɪu/
Optional use of /ŋg/ for certain instances of /ŋ/	'clingy'	/ˈklɪŋ.i/	/ˈklɪŋ.gi/

Figure 4: Some features of New York English

Feature	Example	Am(Gen)	Am(S)
Non-rhoticity	'heart'	/hɑɹt/	/hɑːt/
Use of /ɪ/ rather than /i/ in certain environments	'happy'	/ˈhæ.pi/	/ˈhæ̞.pɪ/

Figure 5: Some features of South Carolina English

3. RULE FRAMEWORK

Previous work had produced a pronunciation lexicon containing over 110,000 words, for use in diphone synthesis. These were transcribed in RP and General American, using machine-readable phonetic alphabets, and parts of speech were also included in the entries. This lexicon was used as the basis for the current work. The RP transcriptions were used as the basic input for generating the British accents, while the General American transcriptions were used to generate the American regional accents.

3.1. Alignment Rules

A number of the rules rely on descriptions of relationships between the original pronunciation and the spelling. For example, part of the rule for producing /x/ in Edinburgh English can be stated as follows:

> Replace a /k/ or /g/ which represents 'ch' or 'gh', and is not part of a syllable-initial cluster, with /x/.

We then need an alignment to distinguish between the /k/ in RP 'lochside', which represents orthographic 'ch' and should be converted to Br(Sc) /x/, and the /k/ in RP 'dockside', which represents orthographic 'ck' and so remains as /k/ in Br(Sc). It is easy to see the correspondence between the orthography and the pronunciation, but less easy to formulate rules to express this accurately (see [2]). An alignment algorithm was designed for the existing lexicons, grouping letters or short sequences of letters with phonemes or sequences of phonemes; the output of this was used as the input to the transformation rules. Syllable boundaries and stress patterns were retained in the alignment as they were often useful for transformation rules.

3.2. Remapping Rules

The first and simplest step in creating regional pronunciations was to remap the correspondences between machine-readable symbols and phonemes for each accent, to allow for different phonemic inventories. These remappings are context-free. In many cases, this allowed the regional accents to use the same machine-readable transcription as the standard accent. For example, Leeds English does not differentiate between /ʊ/ and /ʌ/, whereas RP has both. The symbols 'u' and 'uh', which represent /ʊ/ and /ʌ/ respectively for RP, can both be remapped to represent /ʊ/ in Leeds English. This gives us:

Word	Machine-readable transcription	Br(RP)	Br(N)
'put'	p * u t	/pʊt/	/pʊt/
'putt'	p * uh t	/pʌt/	/pʊt/

Figure 6: Remapping of 'u' and 'uh' for Leeds English

3.3. Rewrite Rules

The second method used, and the most important one, was context-sensitive rewrite rules, based on the existing transcriptions but also permitting other information in the lexicon, such as part of speech, to be used as input. The rewrite rules fall into a number of categories, as described below. Some of the examples have been simplified here due to lack of space.

For some of the rules a number of different formulations would be possible. For instance, glottalisation of /t/ may vary by phonetic environment and social context as well as speaker, with final /t/ being transformed to a glottal stop more readily than medial /t/. For this project, a typical set of environments was used for such cases.

3.3.1. Pre-lexicon Transformations

These are rules for producing a basic pronunciation lexicon for each accent.

a) Obligatory rules - a set of rules which are always applied, for example non-rhoticity in South Carolina:

   'start': Am(Gen) /stɑɹt/ → Am(S) /stɑːt/

b) Obligatory lexical features - isolated words which have unpredictable regional pronunciations, for example 'with' in Edinburgh English:

   'with': Br(RP) /wɪð/ → Br(Sc) /wɪθ/

c) Optional rules - a set of rules which may optionally be applied. These rules give 'broader' pronunciations than the obligatory rules alone, for example, use of /ɪn/ rather than /ɪŋ/ to represent '-ing' in various accents, including Cardiff:

'thinking': Br(RP) /ˈθɪŋk.ɪŋ/ → Br(W) /ˈθɪŋk.ɪn/

d) Optional lexical features - isolated words which for some speakers have unpredictable pronunciations in the regional accent, for example 'make' in Leeds English:

'make': Br(RP) /meɪk/ → Br(N) /mɛk/

3.3.2. Post-lexicon Transformations

These rules apply to the output of the pre-lexicon transformations, and concern allophones, which it is not necessary to include in a lexicon. Some allophone rules were included in the pre-lexicon transformations if they had complex contextual descriptions, including for example morphological information.

The allophones are variants of a single phoneme (though in a few cases, such as Edinburgh /aˑe/ - /ʌi/, it is not clear whether a given alternation is allophonic or phonemic). Allophones are used in all specified contexts, with no lexical exceptions. Some of these would be produced naturally by subjects recording diphones, but others rely on a wide context (for example, in South Carolina vowels may be conditioned by the vowel in the next syllable). Rules are therefore needed to specify the contexts of these allophones.

a) Obligatory - in natural speech, these would be produced by all subjects with the given accent, for example use of taps in various American accents:

'catty': Am(E) /ˈkæ.ti/ → /ˈkæ.ɾi/

b) Optional - in a natural situation these may vary according to the subject or the formality of the situation, for example Edinburgh glottal stops:

'hot': Br(Sc) /hɔt/ → /hɔʔ/

3.3.3. Connected Speech Rules

As some accents have rules which apply in connected speech, these have been included in the framework.

a) Obligatory rules - these include removal of pre-consonantal word-final /r/ in Cardiff English. (This has been transcribed in non-rhotic accents to allow for linking 'r'.)

'car park': Br(W) /kaːɹ paːk/ → /kaː paːk/

b) Optional rules - Leeds English may use /r/ instead of pre-vocalic word-final /t/:

'shut up': Br(N) /ʃʊt ʊp/ → /ʃʊɹ ʊp/

4. RESULTS

The remapping rules cover a fair number of cases, and are straightforward. More interesting issues arise from the rewrite rules.

4.1. Relationship between British and American Pronunciations

Sometimes the most accurate results are obtained by taking a feature of one of the transcribed accents for use in one of the generated accents of the other country. For example, in Br(RP) the ASCII combination |i@| has been used to represent both /ɪə/ (or /iə/) in words such as 'happier', 'topiary', 'fearing' and 'fear'. However, for Cardiff English this needs to be split three ways - /iə/ for 'happier' and 'topiary', /iː/ for 'fearing', and /jɜ/ for 'fear'. Some generalisations can be made about the phonetic environments in which they occur, but a more accurate transformation can be made by including the Am(Gen) transcriptions in the rule environment. We then have, for |i@| preceding orthographic 'r':

Rule i): where Am(Gen) has /iə/, /iɚ/, /jə/ or /i̩ɛ/, transform Br(RP) /ɪə/ to Br(W) /iə/
Examples: 'happier', 'topiary'

Rule ii): in the environment preceding /r/ plus a vowel, where Am(Gen) has /ɪ/ not preceding a geminate /r/, change Br(RP) /ɪə/ to Br(W) /i/
Example: 'fearing'

Rule iii): other cases of Br(RP) /ɪə/ before orthographic 'r' become Br(W) /jɜ/.
Example: 'fear'

No explicit alignment had been produced for matching the Br(RP) and Am(Gen) transcriptions with each other, and they sometimes had different numbers of syllables, or different stress patterns, for example the alignments for 'topiary' were as follows:

Br(RP)	orthog.	t	o	p	ia	r	y	
	phon.	t	*ou	. p	i@	.r	ii	
Am(Gen)	orthog.	t	o	p	i	a	r	y
	phon.	t	*ou	.p	ii	.~e	.r	ii

Figure 7: Alignments between the orthography and the machine-readable phonetic alphabet for 'topiary' in Br(RP) and Am(Gen)

However, nearly all cases were covered by looking for the relevant sequence at the same location in both transcriptions, and if this failed, comparing the previous and following segments.

4.2. One-to-Many Relationships

Some one-to-many relationships, like the Cardiff example described above in 4.1, can be predicted on the basis of information in the lexicon. However, other one-to-many relationships are problematic. For example, both Edinburgh and South Carolina distinguish between 'hoarse' and 'horse', which in RP and General American are homophones. The difference cannot be predicted from the spelling, as there is no consistent correspondence between the different spellings of this set of words and the different vowels, and nor can it be predicted from the part of speech tags. This type of split is the main problem in generating regional pronunciations by rule, as it cannot be resolved except by hand-tagging of individual lexical items, which is not a linguistically satisfactory solution, and is not practical in the current framework.

4.3. Missing Information

Certain features of the various accents are predictable, but rely on information not currently contained in the lexicon.

4.3.1. Morphology

The primary type of missing information is morphological. Some rules for phonemes or allophones depend on morphological boundaries, but these are not explicitly marked in the lexicon. Some of this information can be deduced from the current format, for example by reference to parts of speech, orthography and the phonetic environment, or by lists of affixes. For Edinburgh English we can use the spelling, pronunciation and part of speech to differentiate between the past tense verb 'mooed', which contains a morphological boundary and so has a long vowel, and the noun 'mood', which does not.

'mooed': Br(RP) /mud/ → Br(Sc) /muːd/
'mood': Br(RP) /mud/ → Br(Sc) /mud/

Not all cases, however, are so transparent, particularly compounds. In South Carolina, /nt/ may optionally be reduced to /n/, following a stressed vowel and preceding a vowel ([1], Vol. 3, p. 552). Syllable boundaries are irrelevant, but the /t/ should not be the first syllable of a free morpheme. So, we have:

'winter': Am(Gen) /ˈwɪn.tɚ/ → Am(S) /ˈwɪn.ə/[2]

Unfortunately, the rule as formulated cannot be prevented from applying to compounds such as 'meantime', wrongly giving us:

'meantime': Am(Gen) /ˈmin.taɪm/
→ Am(S) /ˈmin.aɪm/[2]

4.3.2. Other

It has been suggested ([3], p. 162) that lexical items which are 'learned' are less prone to some kinds of casualisation or reduction processes, such as the use of glottal stops for /t/, than more frequent ones. This factor has not been investigated for the current work, but it is possible that word-length might be an approximation to this. More likely is that word frequency in spoken language (not currently included in the lexicon) would provide a basis for distinguishing such groups of words.

More detailed semantic or etymological information would also be of assistance. For example, /ɬ/ in Welsh is only used in native Welsh names or loanwords, and this information is not available in the lexicon.

[2]Other processes would subsequently apply to these strings, such as allophonic adjustments.

5. EVALUATION

Large-scale evaluation of the output was unfortunately not possible due to the lack of comparable work in this area, but native speakers of the accents were consulted where possible and the transcriptions were compared to descriptions and examples in other sources. The rules used seemed to produce acceptable output for the different accents, but some were more successful than others. Particularly problematic were South Carolina, with its large number of allophones, and Edinburgh, which has a very different vowel system from RP.

The discussions with native speakers were invaluable, as this enabled checking of a wider range of examples than are commonly available in the literature. However, it should be noted that native speakers from the same region did not always agree with each other on the lexical incidence of features, or even in some cases on the phonemic inventory. While some regional features have been studied sociolinguistically (for example, see [3]), others have not, making consistency difficult. One solution to this is to model each accent on a single speaker; another is to study several speakers in order to produce an integrated model of the regional variation.

6. CONCLUSIONS

It is possible to develop regional pronunciations by rule from existing standard pronunciations, and most systemic differences can be covered in this way. However, there are certain features, for some accents in particular, which cannot be accurately generated by this method.

7. REFERENCES

[1] Wells, John C. (1982). "Accents of English." Cambridge: Cambridge University Press.

[2] Lawrence, S.G.C, and Kaye, G. (1986). "Alignment of phonemes with their corresponding orthography." *Computer Speech and Language,* Vol. 1, pp. 153-65.

[3] Reid, Euan (1978). "Social and stylistic variation in the speech of children: some evidence from Edinburgh." In: Peter Trudgill (ed.), *Sociolinguistic patterns in British English,* pp. 158-71. London: Edward Arnold.