NEW ENGLISH RULES FOR THE KTH TEXT-TO-SPEECH SYSTEM

Sharon Hunnicutt
Dept. of Speech Comm., Royal Inst. of Technology, Stockholm

Introduction

A set of English rules is presently being written for the speech synthesis system developed at the Royal Institute of Technology (KTH) in Stockholm. This system is constructed to be language-independent. Rules were first written for Swedish,[1] and an English rule system was first presented in 1975.[2] The focus of the current effort has been the development of a more complete set of grapheme-to-phoneme and lexical stress rules. A set of rules to convert expressions involving numbers to words has also been written for the KTH system, and a small lexicon has been added. (1) The material in this paper is an outgrowth of the process of constructing grapheme-to-phoneme and lexical stress rules for the existing formalism of the KTH system. Expressing the rules in this formalism provided the impetus for a study of the constraints and the opportunities presented by this system, and also led to a categorization of rules in terms of special contexts which signal likely exceptions.

The KTH system accepts unrestricted input text, and its first operation is to convert this text to phonemes. This conversion is accomplished either by a small lexicon or by two parallel sets of rules: a set of grapheme-to-phoneme and lexical stress rules, and a set of number-to-phoneme rules. The remainder of the English system contains phonological rules such as devoicing and flapping which are followed by prosodic rules to determine segment durations and fundamental frequency. The segments are expressed as parameters, and synthesized with an OVE III.[5]

An important feature of the KTH system is a special higher-level programming language, the structure of which is similar to that used in generative phonology.[6] The present effort represents the first large-scale attempt to have someone familiar with the rules of another language use this programming language to express their knowledge of these rules.

(1) The author has written a set of grapheme-to-phoneme and lexical stress rules for English, and has worked extensively with various modules in the text-to-speech system developed at the Massachusetts Institute of Technology (MIT). References concerning this work are given in notes 3 and 4 at the end.

The attempt appears to have been successful; the rules were written quickly, and the discipline of the new formalism provided an inspiring perspective on previous work. The categorization of rules mentioned above and some observations about the KTH formalism and the utility of the higher-level programming language are presented below.

Rule Types and Special Contexts

The types of rules needed to predict the grapheme-to-phoneme correspondence in English may be separated into two groups, basic rules, giving the normal pronunciation, and contextually-dependent rules. The KTH system contains approximately 310 grapheme-to-phoneme rules, 50 of which specify the basic, or most frequent, pronunciation of all single vowels and consonants and some consonant clusters and vowel digraphs.

Remaining rules are rather evenly divided into (a)rules for affixes and (b)rules for consonants and consonant clusters, and for vowels and vowel digraphs in special contexts. There are around 130 rules of each of these two types.

There is some question as to whether affixes in general should be recognized and converted by separate rules. Many affixes would be correctly pronounced by the rules for vowels and consonants, would be correctly analyzed by the stress rules, and are not used in any other rule contexts. On the other hand, the morpheme boundaries they define may be useful in syllabification, and it is possible that they signal some prosodic effects such as reduced duration or less F0 excursion.

Special contexts, in which less frequent grapheme-to-phoneme correspondences occur, are seen to be specified by only about a dozen categories. Furthermore, these categories frequently predict special pronunciations for both consonants and vowels. These categories are shown in Figure 1; the same, or similar, contexts for vowels and consonants are found opposite each other. Examples of graphemes receiving correct pronunciation by rules in these categories are also shown.

Most special contexts can be defined in terms of morpheme boundaries. Some contexts express the notion of morph-initial (1) or morph-final (4), while others specify the first (2), last (5,6,10,11) or only (3) consonant(s) or vowel in a morph. Other special contexts can be defined in terms of suffixes (5,8,10). Vocalic inflectional suffixes (5) signal word-final contexts and the end of free roots (6). Two types of "laxing"

VOWELS	CONSONANTS
1) MORPH-INITIAL eulogy, ewe	1) MORPH-INITIAL knee, white, wrote, pneumatic, years
2) FIRST SYLLABLE thiamine, iodine	
3) SINGLE-SYLLABLE MORPH pie, sour, table, raced, striate, vocal, me	
4) WORD-FINAL / MORPH-FINAL bonnie, toe, potato, bake, ma	4) WORD-FINAL / MORPH-FINAL sing, tic, inch, arguable, pariah
5) PRECEDING C - VOCALIC INFLECTIONAL SUFFIX baking, miles, loner, themes	5) PRECEDING VOCALIC INFLECTIONAL SUFFIX wreathes, antiques, apples, acres
6) PRECEDING C- MORPH-FINAL "e" bake, mile, alone, theme	6) PRECEDING C - MORPH-FINAL "e" wreathe, antique, apple, acre, orange, cheese
7) PRECEDING LIQUID / C - LIQUID air, earth, ward, shoulder, doll, or, ...	7) PRECEDING / FOLLOWING A LIQUID quadrille, orle, place
8) IN THE CONTEXT V - C - V_{high} - V alienate, ameliorate, usual	8) IN THE CONTEXT V - C - V_{high} - V revision, dispersion, Russian, racial, ..ation
9) PRECEDING ANOTHER VOWEL Ptolemaic, meteor, myopic, experience	9) PRECEDING ANOTHER CONSONANT chrome, pick, comprehension
10) PRECEDING C - SPECIAL "LAXING" SUFFIXES ratify, utility, conic, edible	
11) OTHERS ..ook, wa.., ..aste, ..ind, ..igh	11) OTHERS -voiced "s" rules

Figure 1 - SPECIAL CONTEXTS

Note: "C" indicates a single consonant, and "V" a single vowel.

suffixes occur (10), and seven suffixes are included in the more general context specified in (8) which is used to signal palatalization of some preceding consonants and the occurence of a long vowel preceding a single consonant in this position. The most prolific exception-generating contexts are those in which a liquid occurs; thirty such rules are included.

Aspects of the Formalism: A Comparison

Most of the differences in the statement of the KTH rules and the MIT rules stem from the type of rule cycle used in grapheme-to-phoneme conversion. Application of the MIT rules is accomplished in three passes: affix removal, consonant conversion in the remnant (assumed to be a monomorphemic root), and conversion of the remainder, i.e., the vowels and affixes. Suffixes are removed by moving inwards from the right word boundary, and other rules are applied by moving from left to right through the word. In each of the passes, the word is scanned, and the appropriate ordered set of rules for that pass is tried until a match in contexts is found.

Application of the KTH rules is accomplished in one pass through the set of rules. If a rule context matches anywhere in the word, moving left to right, the conversion is made, and the next rule context is compared. This method appears to be much more efficient, and does not require the program code needed in the MIT method to direct the various passes with the appropriate set of rules. In fact, no new code was written for the English system at all: the code existent for the Swedish system serves for the English rules as well.

The major difference between the multi-pass method and this one-pass procedure is in the manner of processing and ordering affixes. Recognition and removal of all affixes as a first step in the MIT algorithm corresponds to less than ten rules in the KTH system which recognize vocalic inflectional suffixes and insert a morph boundary marked with the feature "inflectional." The effect of not recognizing all affixes before consonant conversion appears to be rather small: initial consonant clusters after unrecognized prefixes have been observed to be mispronounced in a few cases in the KTH system. However, the opposite effect may be observed in the MIT system: strings incorrectly recognized as prefixes before application of the consonant rules also lead to mistaken pronunciations.

There is a significant difference in the ordering of suffix rules in the two algorithms. Suffixes in the MIT algorithm are recognized first and converted later (in any order). In the one-pass system, however, suffixes must be listed in the order of their probable occurrence from the right-hand side of the word so that their word-final or morph-final position is verified. A short study was undertaken for the purpose of determining the proper order.

There are several other differences in the processing of affixes. Because all consonants are converted before the recognition of most affixes in the KTH algorithm, those consonants in affixes are also converted. The KTH set therefore contains a few rules which are necessary in order to recognize suffixes containing consonants with multiple pronunciations, e.g., the suffix ic in electric or electricity. Suffixes whose final letter may undergo a spelling change are also listed in two rules. The feature of compatability of parts-of-speech in a compound suffix which is found in the MIT algorithm, has not been implemented in the KTH system. This feature is well-developed, but is not frequently needed, and would require additional code and a table of parts of speech for suffixes.

A number of other differences in the two sets of rules are due to the objective of expressing all rules in the KTH system in the higher-level programming language. The most important difference is in the lexical stress rules, which, in the MIT system, are embedded in code. The KTH rules are expressed in the rule language, and are applied using the same formalism as that used for the grapheme-to-phoneme rules. A rule cycle has not been implemented, but the effect of the cycle has, for the most part, been captured in the rules.

Special stress effects due to suffixation are accomplished in two ways. Stress-carrying suffixes are pre-stressed in the suffix rules by noting primary or secondary stress as a feature of the appropriate vowel. This stress may be adjusted later by the stress rules themselves. Suffixes which have no effect on the stress cycle are preceded by a suffix boundary marker with the feature "minus stress cycle." This feature is also assigned to word boundary symbols such as "space," and "period," and becomes part of the right context in many stress rules.

Unlike the MIT system, the KTH rules provide no device with which to retain graphemes after their conversion to

phonemes. The retention of graphemes in the MIT system provides for the specification of either letters or phonemes in both left and right contexts. As a consequence, a substantial subset of rules differ in specification of context. The KTH rules have not yet been tested on a large set of data, but it is believed that this difference gives neither set of rules an advantage worthy of note.

In addition, the KTH programming language allows each phoneme and punctuation mark to be expressed in terms of distinctive features. This type of specification makes the rules more "transparent" than those in the MIT program where variables are used. The facility of specifying optional elements in this programming language has also allowed rules to be expressed more succinctly in several cases.

The experience gained in writing English rules for the KTH system emphasizes the utility of the higher level programming language in which the rules are written. Future development for other languages is very much recommended.

References

1. R. Carlson and B. Granström. 1976. A Text-to-Speech System Based Entirely on Rules. Conf. Record, IEEE International Conf. on Acoustics, Speech, and Signal Processing, pp. 686-688, Philadelphia, Pa.

2. R. Carlson and B. Granström. 1975. Text-to-Speech Conversion by Ordered Rules. Presented at the Eighth International Congress of Phonetic Sciences, Leeds.

3. S. Hunnicutt. 1976. Phonological Rules for a Text-to-Speech System. AJCL Microfiche 57.

4. J. Allen and S. Hunnicutt; R. Carlson and B. Granström. 1979. MITalk-79: The 1979 MIT Text-to-Speech System. Speech Communication Papers, ASA-50, pp. 507-510, The Acoustical Society of America, New York.

5. J. Liljencrants. 1968. The OVE III Speech Synthesizer. IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, pp. 137-140.

6. R. Carlson and B. Granström. 1975. A Phonetically Oriented Programming Language for Rule Description of Speech. Speech Communication, Vol. 2, Almqvist & Wiksell, Stockholm.