Proceedings of The Institute of Acoustics

The Automated Phonetic Transcription of English Text

S G C Lawrence. B J Williams. G Kaye

Speech Research. IBM UK Scientific Centre.
Athelstan House. St Clement Street. WINCHESTER. SO23 9DR

Introduction

The present work is part of a larger project to produce, from machine-readable text, acceptable and intelligible synthetic speech which can be listened to for extended periods without causing fatigue.

A typical use of such a system is in the correction and proofing of documents. Lee[1] has shown that the ear is more efficient at detecting mistakes than is the eye. For this to be exploited it is essential that any irregularity in the pronunciation of the synthetic speech be the result of an error in the input text, rather than an artifact of the system.

To realise this objective an accurate phonetic transcription must be generated from the input text. This transcription should be based on a well-defined and acceptable model of English pronunciation (RP has been chosen in the present work), and should represent how the text might be read by a competent and careful reader; for example, a BBC news reader. A system has been developed to meet this objective. The present implementation has three stages which produce:-

A phonemic transcription.
An allophonic transcription.
A phonetic transcription with some consonantal assimilation.

Each stage is driven by a set of rules stored apart from the program code. The present version is flexible in its design and implementation. Thus experimental changes may be made to the rules, and the effect checked by the user. It has been developed to run in real time on an IBM Personal Computer, and to display the results using a modified character set that includes the International Phonetic Alphabet and the more important diacritics.

Overall approach

The sequential stages in the transcription are:-

1. Preprocessing to convert numerics to text and to analyse abbreviations to determine if they are to be spelled out letter-by-letter (e.g. UN), or pronounced as a word (e.g. UNICEF).
2. Text-to-phonemic transcription. This is driven by the text-to-phoneme rules, and a table of suffixes.
3. Syllabification of the phoneme string.
4. Phoneme-to-allophone convertion. This is driven by allophone rules.
5. Assimilation of allophones. This is driven by assimilation rules.

At present, this system takes no account of lexical stress. All vowel reduction is embodied within the transcription rules. Such an approach has obvious pitfalls, leading at times to incorrect vowel reduction, and lack of compactness of the rules. A new system is being developed exploiting the techniques of lexical stress assignment developed by Williams [2]. This system will not require the text-to-phoneme rules to account for vowel reduction, which will be conditioned by the lexical stress rules. Another restriction is that no attempt has been made to resolve the pronunciation of homographs. For example, the pronunciation of the word *house* depends on whether it is a noun

/haʊs/ or a verb /haʊz/. This requires the assignment of syntactic categories which is to be handled by the new version of the system.

The Rules

Several spelling-to-sound rule systems have been published, notably those of Ainsworth [3], McIlroy [4], and Elovitz et al.[5]. These all aim to produce some form of phonemic transcription. None of them produce an allophonic transcription, nor do they take account of consonantal assimilation. The present work follows a strategy similar to that of Elovitz, with extensive modifications and additions to handle the allophonic transcription and the assimilations.

The rules for each of the three stages have the same basic format:-

LEFT-C(object)RIGHT-C = transcription

The string being processed is scanned left-to-right. A match is sought between one or more characters in the input string and the **object** section of the rules. When such a match is found the left and right contexts (LEFT-C and RIGHT-C) are checked in turn. If all three match then the contents of **transcription** are added to the output string. The search continues with the characters following **object** in the input string until the whole of the input has been transcribed.

Phonemic Transcription Rules

The following is an example of some of the rules for the letter *e*. They are largely the rules associated with the transcription of *et*, and the endings *se* and *ed*.

RULE	OUTPUT	EXAMPLE
C1M(e)g,e	= ɪ	collEge
(easy)#	= i:zɪ	quEASY
[/k,g,tʃ,dʒ/](et)C0M,#	= ɪt	budgET
C1,C1(et)C0M,#	= ɛt	regrET
[/z,tʃ,ʃ,ʒ,dʒ,s/](es)#	= ɪz	matchES
[/p,t,k,f,θ/](es)#	= s	flukES
V1,C0M(es)#	= z	gamES
V1,C0M,[t,d](ed)#	= ɪd	datED
[/p,k,f,ʃ,s,θ/](ed)#	= t	dupED
V1,C0M(ed)#	= d	lovED
V1(e)#	=	havE
(e)C1,i	= i:	Evict
(e)n,C1	= ɛ	fEnce
(e)	= ɪ	bEtween

The rules are grouped alphabetically by the initial letter of the **object** string and listed in a critical order with a default last rule for each letter, which acts as a catch-all. For example, the rule

i (a) b,l = ə

produces the schwa in the word *reliable* /rɪlaɪəbl/.

If a match with a rule fails, the next rule is tried. Repeated failures lead to the last rule for the given letter, which always produces a successful match. The last rule for the letter *a* is

(a) = æ

Sample output from the system

The following is an example of the output produced by the transcription system.

Original text
The very first showroom is used by young mums with handbags

Broad phonemic transcription
ðə verɪ fɜ:st ʃəuru:m ɪz ju:zd baɪ jʌŋ mʌmz wɪð hændbægz

Syllabification
ðə ve-rɪ fɜ:st ʃəu-ru:m ɪz ju:zd baɪ jʌŋ mʌmz wɪð hænd-bægz

Allophonic transcription
ðəverɪfɜ:st?ʃəʊɾu͡:mɪzju:zd̚baɪjʌ̃ŋmʌ̃mzwɪ̯ðhæ̃nd̚bægz̥

Allophone assimilation
ðəverɪfɜ:ʃʃəʊɾu͡:mɪʒu:zbaɪjʌ̃ŋmʌ̃mzwɪ̯ðhæ̃mbægz̥

Checking the rules

A dictionary of head-words with their pronunciation and part of speech has been abstracted from Collins' English dictionary. It has been used to tag with their phonemic transcription all the words occurring in the LOB corpus [8]. This tagged corpus will be used to check the accuracy of the rules against the most frequently-used words in English.

References

[1] Lee, W.R. (1957) Spelling Irregularity and Reading Difficulty in English. Slough: National Foundation for Educational Research. Occasional Paper No. 2.

[2] Williams, B.J. (1984) Lexical stress assignment in English; a metrical approach.

[3] Ainsworth, W.A. (1973) A System for Converting English Text into Speech. IEEE Transactions on Audio and Electroacoustiics, Vol. AU-21, No 3, June 1973.

[4] McIlroy, M.D. (1974) Synthetic English Speech by Rule. Computing Science Technical Report #14, Bell Laboratories, March 1974 (revised Sepetember 1974).

[5] Elovitz, H.S., Johnson, R.W., McHugh, A., Shore, J.E. (1976) Automatic Translation of English Text to Phonetics by Means of Letter-to-Sound Rules. U.S. Naval Research Laboratory Report 7948, January 1976.

[6] Ladefoged, P. (1975) A Course in Phonetics. Harcourt Brace Jovanovich, Inc.

[7] Wells, J.C., Colson, G. (1971) Practical Phonetics. Pitman.

[8] LOB (Lancaster-Oslo/Bergen) Corpus (1978) Department of English, University of Oslo

Acknowledgements

It is a pleasure to thank William Collins Sons and Co. for making available for our use a copy of the Collins English Dictionary in computer-readable form.