Problems Relating to the Phonetic Encoding of Words in the Creation of a Phonetic Spelling Recognition Program

Michael Higgins and Wang Shudong

Department of Kansei Design Engineering,
Faculty of Engineering and Technology, Yamaguchi University
2-16-1 Tokiwadai, Ube City, Yamaguchi Prefecture, Japan
{higginsm, peterwsd}@yamaguchi-u.ac.jp

Abstract. A relatively new area of research in centering on the phonetic encoding of information. This paper deals with the possible computer applications of the Sound Approach© English phonetic alphabet. The authors review some preliminary research into a few of the more promising approaches to the application of the processes of machine learning to this phonetic alphabet for computer spell-checking, computer speech recognition etc. Applying mathematical approaches to the development of a data-based phonetic spelling recognizer, and speech recognition technology used for language pronunciation training in which the speech recognizer allows a large margin of pronunciation accuracy, the authors delineate the parameters of the current research, and point the direction of both the continuation of the current project and future studies.

1 Introduction

In 1993-1994, Dr. Michael Higgins of Yamaguchi University, Japan developed and did initial testing on a new system of phonetic spelling of the sounds in English as an aid to learning better English pronunciation and improving listening and spelling skills in English for Japanese students of English. The method, subsequently entitled "A Sound Approach", has been proven to be a very effective English phonetic system [1]. The Sound Approach (SA) alphabet represents without ambiguity all sounds appearing in the pronunciation of English language words, and does so without using any special or unusual symbols or diacritical marks; SA only uses normal English letters that can be found on any keyboard but arranges them so that consistent combinations of letters always represent the same sound, for example, for the word, "this", instead of using IPA (International Phonetic Alphabet) symbols of /ðis/, SA uses /dhis/ to express the pronunciation. Consequently, any spoken word can be uniquely expressed as a sequence of SA alphabet symbols, and pronounced properly when being read by a reader knowing the SA alphabet. For instance, the sentence "One of the biggest problems with English is the lack of consistency in spelling," is written in SA (showing word stress) as: "Wun uv dhu BI-gust PRAA-blumz widh EN-glish iz dhu lak uv kun-SIS-tun-see in SPEL-ing. "

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613, pp. 1256 – 1260, 2005.

2 Project Development

Due to representational ambiguity and the insufficiency of English language characters to adequately and efficiently portray their sounds phonetically, the relationship between a word expressed in SA alphabet and its possible spellings is one to many. That is, each SA sequence of characters can be associated with a number of possible, homophonic sequences of English language characters (e.g. "tuu" is equivalent to "to", "too", and "two"). However, within a sentence usually only one spelling for a spoken word is possible. The major challenge in this context is the recognition of the proper spelling of a homophone/homonym given in SA language.

In addition to the obvious speech recognition OS that would eventually follow, automated recognition of the spelling has the potential for development of SA-based phonetic text editors which would not require the user to know the spelling rules for the language but only being able to pronounce a word within a relatively generous margin of error and to express it in the simple phonetic SA-based form. Speech recognition parameters could also be adjusted to accommodate the wider margin of error inherent in SA which is based on International Broadcast Standard English. As the SA is phonetic (i.e., a clear one-to-one correlation of sound to spelling), the words themselves could be 'regionally tuned' so that one could effectively select the regional accent that they are accustomed to much in the same way that we currently select the keyboard by language groupings. In other words, someone from Australia could select Australia as their text editor and phonetically spell the word 'day' as 'dai' without contextual ambiguity. (See Fig. 1)

I'd like to go to the hospital again today.
1) Ai'd laik tuu go tuu dhu HAAS-pi-tul u-GIN tuu-DEI. (International Standard Broadcast English) 2)Aa'd lak tu go tu dhu HAAS-pi-tul u-GEE-un tu-DEI. (Generic Southern US English) 3) Oi'd laik tuu go tuu dhu HOS-pi-tul u-GEIN tuu-DAI. (Australian-English)

Fig. 1. Regionally tuned sample sentences

The approach adapted in this project involves the application of rough sets [2] in the development of a data-based word spelling recognizer. In this part, the techniques of rough sets, supported by rough-set based analytical software such as KDD-R [3], would be used in the analysis of the classificatory adequacy of the decision tables, and their minimization and extraction of classification (decision) rules to be used in the spelling recognition. While the initial identification and minimization of the required number of information inputs in such decision tables would be one of the more labor intensive aspects of the project, it should be emphasized at this point that the latter stages of the process of minimization and rule extraction would be automated to a large degree and adaptive in the sense that inclusion of new spoken word-context combinations would result in regeneration of the classification rules without human intervention. In this sense the system would have some automated learning ability allowing for continuous expansion as more and more experience is accumulated while being used [4]. The adaptive pattern classification part of the system development is absolutely key to the successful deployment of the system. This aspect of the system, however, only becomes important when some of the more essential problems are solved. Given that, let us briefly outline how we plan to use rough sets in our approach.

In the word spelling recognition problem, in addition to the problem of homophones, one of the difficulties is the fact that many spoken words given in SA form correspond to a number of English language words given in a standard alphabet. To resolve, or to reduce this ambiguity, the context information must be taken into account. That is, the recognition procedure should involve words possibly appearing before, and almost certainly after the word to be translated into Standard English orthography. In the rough-set approach this will require the construction of a decision table for each word. In the decision table, the possible information inputs would include context words surrounding the given word and other information such as the position of the word in the sentence, and so on. (See Appendix)

Several extensions of the original rough sets theory have been proposed recently to better handle probabilistic information occurring in empirical data, and in particular the Variable Precision Rough Sets (VPRS) model [4] which serves as a basis of the software system KDD-R [3] to be used in this project.

In the preliminary testing of SA completed in 1997, a selection of homonyms was put into representative sentences. The words in the sentences were assigned numbers (features) according to a simple, and relatively unrefined, grammatical protocol. These numbers were then inserted into decision tables and using KDD-R it was found that the computer could accurately choose the correct spelling of non-dependent homonyms (i.e., those homonyms for which the simple grammatical protocol was unable to determine the correct spelling from the context) 83.3% of the time, as in the sentence, “The ayes/eyes have it.” With dependent homonyms, as in the sentence, “We ate eight meals,” the computer could accurately choose the correct spelling more than 98% of the time [4].

Besides the above usages of SA, we are also testing to build HMM/GMM modules of speech recognizers for English pronunciation training based on SA for an online English pronunciation training system, currently, just for Japanese learners of English. As SA encoded speech recognizer allows larger scope of parameter baseline limitation than any current existing speech recognizer, it would more accurately catch the real error of English pronunciation. For example, many acceptable /f/ pronunciations in Japanese English words are judged as wrong or unable to recognize in the current English pronunciation training system using ASR technology [5].

The “regionally tuned” feature of SA can be effectively used for foreign language pronunciation training, too. Usually a non-native speaker’s English pronunciation is a mixture of accents of American English, British English, Australian English, etc, and of course colored by his/her own mother tongue. In this case, if ASR modules are built based on SA, then users’ input sounds will be encoded to SA with a large margin of speech signal parameters. From SA codes, spoken words are re-coded to regular text on the basis of one-to-one correspondence of SA to regular word, the speech recognition rates will be greatly improved accordingly. Additionally the ASR software training time will be much shorter. Therefore, the speech of foreign language learners who read continuous sentences into a computer will be more easily recognized and translated into correct texts (STT). Also the English learner does not need to worry about what kind of English he/she has to speak.

3 Current Challenges

The adaptive pattern classification part of the system development presents one of the largest difficulties. As alluded to above, a major difficult with the approach suggested is that the practicality of re-coding every word in the English language into a Sound Approach coding is problematic, simply due to the size of the problem, and the fact that it is difficult, at this time, to accommodate any practical public machine learning, whereby users could add new words to the Sound Approach© coded dictionary, which is the way normal spell check systems overcome the problem of an initially small dictionary size. To ease this problem, an already phonetically coded dictionary using the IPA symbols as pronunciation guides has been temporarily adopted, and an interface to link the dictionary (IPA) coding with the Sound Approach coding is being developed. This obviously will save a lot of time, as so many words have already been encoded. However, the problem is that such dictionaries do not contain the phonetic coding of all the inflections of a word, e.g. go = goes = went, or play = playing = played = plays. Therefore, we are still faced with the problem of having to encode many words by hand, before the system can be used as a practical phonetic spell checker. This problem must be solved before we can seriously consider other issues such as how to process words ‘in context’ and so on.

4 Prospects and Conclusion

If these particular problems can be adequately addressed in the coming year, the authors see no major difficulty for being able to complete the classification of the homonyms into decision tables and, using that as a base, develop a protocol for converting words written in Standard English or IPA symbols into SA characters for ease of encoding. In this way, the interactive database of SA to Standard or Standard to SA could be completed in a relatively short amount of time. This will then make completing the spelling recognition project possible with a more complete regional tunability and, in turn, pave the way to complete voice recognition capability.

References

1. Higgins, M.L, with Higgins M.L and Shima, Y.: Basic Training in Pronunciation and Phonics: A Sound Approach, vol. 19, number 4, The Language Teacher (1995) 4-8.
2. Ziarko,W.: Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer Verlag (1994)
3. Ziarko,W and Shan, N.: KDD-R: A Comprehensive System for Knowledge Discovery Using Rough Sets. Proceedings of the International Workshop on Rough Sets and Soft Computing, San Jose (1994) 164-173.
4. Higgins.M.L, with Ziarko, W.: Computerized Spelling Recognition of Words Expressed in the Sound Approach. New Directions in Rough Sets, Data Mining, and Granular-Soft Computing: Proceedings, 7th International Workshop, RSFDGrC '99. Lecture Notes in Artificial Intelligence 1711, Springer Tokyo (1999). 543-550
5. Goh Kawai and Keikichi Hirose: A Call system for teaching the duration and phone quality of Japanese Tokushuhaku Proceedings of the Joint Conference of the ICA (International Conference on Acoustics) and ASA (Acoustical Society of America) (1998) 2981-2984

Appendix

Values of the observations
(Grammatical Protocol):
0: none 1.verb
2: noun/pronoun 3: adjective
4: adverb 5: article
6: connective 7: number
8: possessive a: let, please, etc.
b: will, shall, can (modals), etc.
c: prepositions

Fig. 2. Values

Head	-2	-1	Spelling
1	2	c	aye
2	2	1	aye
3	0	0	aye
4	0	5	ayes
5	1	8	eye
6	2	5	eye
7	0	8	eyes
8	0	5	eyes
9	1	1	eyes
10	1	5	eye
11	0	2	eyed
12	5	2	i
13	0	0	I

Non-dependent

Fig. 3. Sample Table 1b — Reduct: "ai"

Table 1. Sample Table 1a: "ai" (IPA symbol: aI); (Sound Spelling: ai)

Head Word	Sentence Number	-5	-4	-3	-2	-1	Spelling
1	15	2	b	1	2	c	aye
2	16	7	2	c	2	1	aye
3	17	0	0	0	0	0	aye
4	18	0	0	0	0	5	ayes
5	19	0	0	2	1	8	eye
6	20	0	2	1	2	5	eye
7	21	0	0	0	0	8	eyes
8	22	0	0	0	0	5	eyes
9	23	0	0	2	1	1	eyes
10	24	1	1	c	1	5	eye
11	25	0	0	0	0	2	eyed
12	26	0	0	0	5	2	i
13	27	0	0	0	0	0	I

Sample Sentences:

15. "I'll love you for aye."
16. "All those in favor say, 'aye'."
17. "Aye, Captain."
18. "The 'ayes' have it."
19. "He injured his eye at work."
20. "He gave me the eye."
21. "Her eyes are blue."
22. "The eyes have it."
23. "She's making eyes at me."
24. "I'm going to keep an eye on you."
25. "He eyed the situation carefully before he went in."
26. "The letter i comes after the letter h and before j."

27 "I want to go out tonight."