A COMPARISON OF SEVERAL COARSE PHONETIC CLASSIFICATION SCHEMES —— PRELIMINARY RESULTS

Edmund M.-K. Lai BE & Y. Attikiouzel BSc, PhD

Department of Electrical & Electronic Engineering
University of Western Australia

ABSTRACT — The performances of three different coarse phonetic classification schemes for a very large lexicon are measured in terms of the number of unique cohorts, average cohort size and expected cohort size. It was found that the seven-way classification proposed by the authors is good for relatively short words with 3 or 4 phonemes.

INTRODUCTION

There is increasing evidence that the problem encountered in large-vocabulary as well as continuous speech recognition could only be solved by applying syntactic and semantic knowledge of speech at various levels (Moore,1985). In the area of isolated-word large-vocabulary recognition, Shipman and Zue(1982) reported that there are strong phonotactic constraints on English words. They suggested a means of lexical access by segmenting an utterance into a sequence of coarse phonetic classes. In this way it is possible to restrict the search for the correct word to a small subset of the lexicon. They have shown that almost one third of the entries in a 20,000-word lexicon can be uniquely identified by a sequence comprising of coarse phonetic classes alone. Understanding of the phonotactic constraints is also important to phoneme-based speech recognition that does not require segmentation such as the method proposed by Kashyap (1979).

A number of coarse phonetic classification schemes has been proposed. Shipman and Zue (1982) suggested a scheme of classification of phonemes into six coarse classes. These six classes are vowels, stops, nasals, strong fricatives, weak fricatives and liquids/glides. Carlson et. al. (1986) performed a detailed statistical analysis of the phonetic properties of five European languages using this six-way classification. Lagger and Waibel (1985) proposed an alternative classification scheme which includes the six coarse classes: fricatives, stops, nasals, front vowels, back vowels and /r/. Derouault et. al.(1986) recently suggested three other six-way classifications based on French.

The purpose of this study is to compare the performance of three classification schemes. Two of the these are proposed previously by other researchers while the third one is proposed by the authors. The new scheme consists of seven coarse phonetic classes of which three are vowels. An initial analysis on the performance of the three schemes in the case where the word-final stop consonant is ignored are also compared. This provides an indication of the each scheme's tolerance to missing word-final consonant information.

DATA PREPARATION

The lexicon that is used in this study is extracted from the Macquarie Dictionary. It includes all the single words that appear in the main entry of the dictionary. Entries that contain more than one word are excluded. Suffixes, prefixes and entries which are not complete words are also excluded. This results in a lexicon with a size of 51018 words.

There are words with more than one possible pronounciations. In such a case the first one that appear in the pronounciation field of that entry in the dictionary will be selected.

COARSE CLASSIFICATION SCHEMES AND ANALYSIS PROCEDURE

Three different coarse phonetic classification schemes are studied. The first one is the Shipman and Zue (1982) classification (C1). The second one is the Lagger and Waibel (1985) classification (C2). The last one is one proposed by the authors (C3) that classifies the phonemes into 7 classes: Front vowels, Central vowels, Back vowels, Nasals, Plosives or Stops, Fricatives and Liquids/Glides.

C3 is designed to match the ratio of the number of consonant classes to that of vowels close to the actual consonant-vowel ratio of the lexicon.

RESULTS AND OBSERVATIONS

General Statistics

A total of 377653 phonemes appear in the 51018+word lexicon and the average number of phonemes per word is 7.4. This average is higher than that obtained by Carlson et. al. (1985) — 5.96, and that reported by Shipman and Zue (1982) — approximately 6.5 for the 20,000+word lexicon. This is due to the fact that the lexicon used in the present study includes a significant number of less frequently used words which are generally phonetically longer compared with the more frequently used ones.

Figure 1 shows the number of words in the lexicon with specific phonetic lengths. The longest word contains 24 phonemes. 43.95% of the phonemes are vowels while the remaining 56.05% are consonants. In general it was found that the vowel content is richer in long words than in the short ones.

Normal Cohort Statistics

Figure 2a shows the percentage of unique cohorts among cohorts with the same phonetic length. Since the number of unique cohorts can be taken as a measure of the relative difficulty of speech recognition using the particular classification scheme, it is observed that scheme C3 performs better than the other two especially for the shorter words which are more common.

The other two measures are the average cohort size and the expected cohort size as define by Waibel (1982). A comparison of the average and expected cohort sizes for the three schemes are shown in figures 3 and 4 respectively. A breakdown of the two measures of words with various phonetic lengths is presented in figures 5a and 6a. Scheme C3 again performs better than the other two for shorter words. Note that the performance of C2 is much closer to that of C3 than C1.

Cohort Statistics with Missing Stop Consonant

Figure 2b shows the percentage of unique cohorts and figures 5b and 6b show the average and expected cohort sizes for words of different phonetic length respectively. The classification schemes C1, C2 and C3 are identified as C1a, C2a and C3a respectively in these figures. Under the circumstance where the word-final stop consonant is ignored, the performance of all three classifications degraded as expected. However, the degradation is worst for C1, followed by C2.

CONCLUSIONS

This study is part of our on+going research in very large vocabulary speech recognition. We have presented the cohort statistics of the three classification schemes under the normal circumstances as well as in the case where the word-final consonant is ignored. From the results we can expect that a seven-way classification will ease the recognition task for words with 3 or 4 phonemes considerably. However, the performance of the Lagger and Waibel (1985) classification is much closer to our seven-way classification which indicates that the ratio of consonant to vowel classes are important as it should reflect the actual consonant-vowel ratio of the lexicon. Since 3 out of the 7 classes of the new classification are vowels, the actual classification task is also easier compared with the 1 out of 6 (C1) or 2 out of 6 (C2) ratio for vowel classes.

ACKNOWLEDGEMENT

The authors wish to thank Macquarie Library Pty. Ltd. for the permission to use the Macquarie Dictionary and the Speech and Language Research Center of Macquarie University for providing the machine+readable version of the dictionary for this research.

REFERENCES

CARLSON,R., ELENIUS,K., GRANSTROM,B. & HUNNICUTT,S. (1985)
"Phonetic and Orthographic Properties of the Basic Vocabulary of Five European Languages", Speech Transmission Lab. Quaterly Progress Report, STL-QPSR 1/85, pp.63-94.

CARLSON,R., ELENIUS,K., GRANSTROM,B. & HUNNICUTT,S. (1986) "Phonetic Properties of the Basic Vocabulary of Five European Languages: Implications for Speech Recognition", Proceedings of ICASSP '86, pp.2763-2766.

DEROUAULT,A.+M., MERIALDO,B. & SOUDOPLATOFF,S. (1986) "Phoneme Classification using Markov Models", Proceedings of ICASSP '86, pp.2759-2762.

KASHYAP,R.L. (1979) "Syntactic Decision Rules for Recognition of Spoken Words and Phrases Using a Stochastic Automaton", IEEE Trans. Pattern Anal. & Machine Intelligence, Vol.PAMI-1, No.2, pp.154-163.

LAGGER,H. & WAIBEL,A. (1985) "A Coarse Phonetic Knowledge Source for Template Independent Large Vocabulary Word Recognition", Proceedings of ICASSP '85.

MOORE,R.K. (1985) "Systems for Isolated and Connected Word Recognition", in New Systems and Architectures for Automatic Speech Recognition and Synthesis, R. De Mori & C.Y. Suen (ed.), Springer-Verlag.

SHIPMAN,D.W. & ZUE,V.W. (1982) "Properties of Large Lexicons: Implications for Advanced Isolated Word Recognition Systems", Proceedings of ICASSP '82, pp.546-549.

WAIBEL,A. (1982) "Towards very large vocabulary word recognition", Report CMU-CS-82-144.

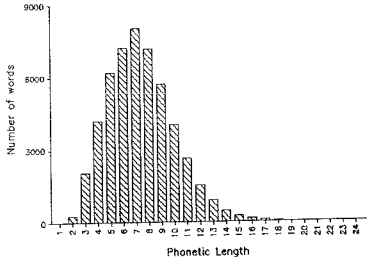

Figure 1: Number of words with different phonetic lengths

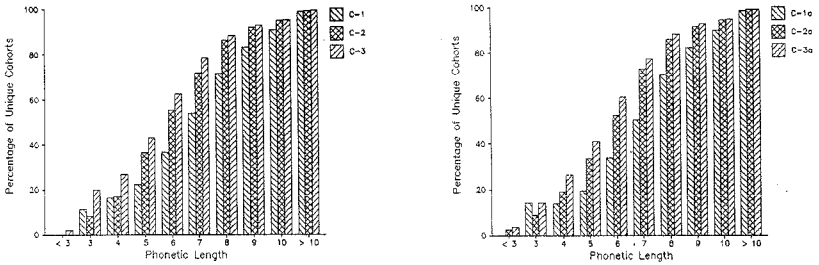

Figure 2: Amount of unique cohorts with different phonetic lengths

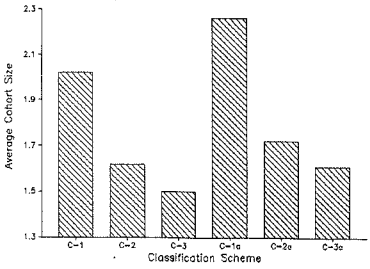

Figure 3: Average cohort size of each classification

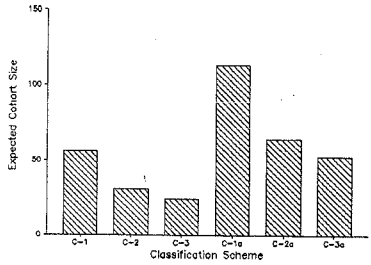

Figure 4: Expected cohort size of each classification

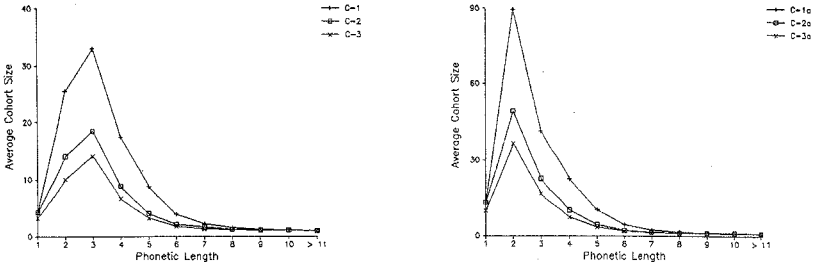

Figure 5: Average cohort size of words with different phonetic lengths

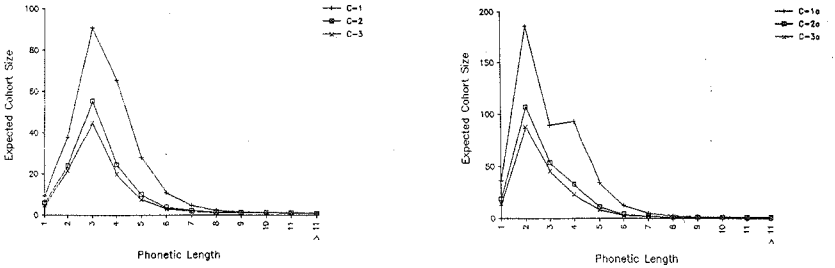

Figure 6: Expected cohort size of words with different phonetic lengths