The construction of phonemic knowledge using Clustering Methodology

Hisham Darjazini and Dr Jo Tibbitts

Department of Electrical Engineering, University of Western Sydney Nepean

ABSTRACT - In order to simulate inherent human phonemic knowledge, a mechanism for categorising phonemic information and for providing efficient structures to refine information is proposed. A model is used that depends on the hierarchal structure of the knowledge which classes the phonemic information in clustering manner. The classification hierarchy takes advantage of inheritance where once the phoneme is identified within the phonemic stream, it inherits knowledge of all the statistical possibilities of the remain stream. It is suggested that this process will enhance isolated word recognition as well as continuous speech recognition.

SUMMARY

This paper describes the preliminary work that has been done to identify the statistical order of phonemes in initial and subsequent placing. This statistical information will be presented to a neural network based architecture to assist the network in phoneme identification by taking into account some global features such as likelihood of occurrence of one phonetic unit given a previous history of units. By including this statistical feedback into the neural net, it is suggested that the recognition task has the potential to be more efficient and to provide statistical reinforcement when acoustic cues are insufficient or noise is present.

INTRODUCTION

Phonemic Knowledge Representation

The knowledge required to recognise speech is derived from raw data which is refined, processed and/or analysed to yield information. This information is converted into phonemic knowledge with further processing and the addition of heuristic information. A phonemic knowledge system simulates the knowledge that is gained in learning a language. It consists of two essential elements: data items represents phonemes in Australian English and interpretive procedures represent a technique of combining these data items together. As data is only meaningful in its context, data structures without interpretive procedures are not useful.

In large vocabulary speech recognition systems, it is necessary to develop a highly reliable phoneme recognisor (Katsuhiko et al. 1990) to minimise the number of final errors made. There have been several successful phoneme recognisors using Hidden Markov Models and/or Artificial Neural Networks (ANN). ANNs have attracted considerable interest in recent years due to their performance in the speech recognition task. ANNs have been successfully applied to obtain high performance phoneme recognition. Large vocabulary recognition requires sequential control of subword units such as syllables or phonemes. It is suggested that an efficient large vocabulary recognition system can be constructed using neural nets as subrecognisors that are controlled by phonemic knowledge. (Hanazawa et al. 1989, McDermott et al. 1989, and Sawai et al. 1990)

According to Feigenbaum (1989), the phonemic knowledge system has to satisfy the following three requirements. It has to be able to encode information about the properties of the data items. It needs to have procedures on how to perform certain tasks (ie an algorithm). It also needs to have a cognitive process embedded in the procedures that improves the reliability and evaluates the performance of the system. There is the potential for upgrading the initial structure within the confines of staying within the boundaries of the initial knowledge.

The stages involved in the use of the phonemic knowledge are acquisition and retrieval. Acquisition

involves the integration of new information into the knowledge system which occurs at two levels. The lower level is concerned with structuring facts in a data base. The higher level is called *learning* which integrates new information within existing knowledge structures. The retrieval stage is called *recall*. Retrieval requires sorting through the knowledge base for the particular facts. This behaviour is achieved by using the concepts of linking and lumping; well known in Artificial Intelligence literature. Linking extracts information that is implied from the raw data. Lumping combines component structures into a larger structure.

Clusters and Hierarchical Structures

Categorisation of the phonemic information into phonemic groups provides an efficient structure for refining the information and organising it as phonemic knowledge. Classification involves the categorisation into phonemic groups, classes, subclasses, and orders. Every group contains set of classes, the classes may be classified within this hierarchical structure. The advantage of the classification hierarchy is the concept of inheritance which specifies that once it has been classified, the phonetic class automatically inherits all classes, subclasses of that classification. The hierarchical structure of the phonemic knowledge has two levels of hierarchies. The base level is the phonemic information. The upper level is hierarchy generated by combining phonemic information in clusters.

Table 1 shows how textual information can be categorised into groups containing one or more phonemes. The group for the letter, A, for example contains six phonemes whereas the group for the letter K contains two phonemes. This group then specifies the phonemes that could be appropriate given the occurrence of that letter within text.

Figure 1 illustrates the distribution of the phonemic data into phonemic groups, classes, and subclasses. The whole units here distributed into clusters which form hierarchy structure. The connection through the hierarchy is weighted by the conditional probability of the class or subclass given a group, class or subclass.

To accelerate the recognition cycle, the phonemic knowledge is developed in two parts. Both parts have adopted the same knowledge construction methodology that depends on hierarchical clustering of the phoneme data. This methodology is used in many artificial intelligence applications and plays vital role in different expert systems (Tam, 1993). It is adopted here to enhance the ANN performance. The concept of the knowledge representation is taken from this methodology and reflected in the behaviour of the neural network system to make its responses to a specific stimuli similar to those responses of human knowledge. Within this working paper, we describe the mean of categorising the phonemic information to provide an efficient structure for refining information and organising it as a knowledge base. We also describe the preliminary architecture of the ANN that utilises this knowledge base.

DESCRIPTION OF MODEL

Figure 2 shows an block diagram of the model of the speech recognition system. The phonemic knowledge system being discussed in this paper is shown to contain a Knowledge Processor and a Memory System. The Knowledge Processor runs the procedures of learning and recalling the phonemic data within the limits of Feigenbaum's requirements. The Memory System stores the basic phonemic units in terms of vectors and also contains the conditional probability distribution for all phonemes in the initial and any sequential position. Speech in the form of vectors is fed into the neural network along with feedback on statistical likelihood of the incoming phoneme in the form of a conditional probability. The output from the network is a vector representing the classification of phonemes. This vector is input to the Knowledge Processor as feedback in the recall operation. The decoder converts this output speech vector into ASCII format. The Accumulator is a FIFO memory system that gathers sequential data output.

In this model, a set of 1410 words are used to build up the initial phonemic statistical structure. These words were extracted from three different contexts which were an Engineering report, a financial report and general article. Each transcription included some conversational speech. The words were converted to phonemic representation based on the International Phonemic Alphabet used in Australian English (Macquarie Dictionary).

The phonemic structure of the word set is used as basis for representation of the initial phonemic knowledge. The statistical nature of the knowledge will be represented by the conditional probability for that specific phoneme relating to the previous history of recall cycles. This constructed knowledge will be used as activating input to ANN, and controls the whole behaviour of the ANN. The Phonemic Knowledge system acts with sub-recognisors as a Large Vocabulary Recognition System. It is suggested that including this additional statistical information will enhance the performance of the speech recognition system especially when the acoustic cues presented at the ANN inputs are insufficient or there is high level of noise associated with speech signal.

The first level of the phonemic knowledge is the phonemic units themselves represented by raw data collected from phonemes in Australian English. This level contains 45 different phonemes taken as blocks to be represented in the phonemic knowledge under construction, with the addition of an extra block to represent silence (see figure 1b). These blocks are encoded as output states of the ANN, while simultaneously being feedback to the inputs of the neural net elements, (see Figure 2). The overall behaviour of the network is affected by these additional inputs in three ways. Firstly they offer knowledge of data during the recognition or recall cycle. Secondly they update clusters during the learning cycle of the system. Thirdly they detect possible mistakes in the network behaviour.

The second level of the knowledge is hierarchy clusters constructed on the basis of the conditional probabilities of occurrence for phonemes depending on previous stream of phonemes. These two levels have been inferred from classifying the phonetical data which was derived for the 1410 words. The knowledge hierarchy is also divided into two parts. The first is for starting the system; the second is for the continue the recognition cycles. The conditional probabilities of phonemes for the starting and continuous parts have been extracted and represented in the knowledge clusters.

RESULTS, CONCLUSION AND DISCUSSION

When starting the system, thirteen different groups were observed which are {T, A, H, O, I, W, S, P, E, B, F and D}. Each of these groups has a different probability of occurrence. Figure 3 shows the likelihood of each phoneme group occurring in the starting position. This likelihood is adopted to order and distribute the classes and subclasses.

When continuing the recognition cycle, twenty three different groups were observed which are {T, A, O, I, H, C, W, S, P, F, E, B, D, M, N, G, R, L, Y, U, K, J and V}. Figure 4 shows the likelihood of occurrence of each group in the starting position. The correct identification of the starting position is used as the basis for sequential searching and identification procedures. Both parts have the same hierarchical structure with clustering phonemes according to conditional probability of their occurrence. The spatial separation of the knowledge into two parts is intended to accelerate both learning and recalling cycles, and satisfies Feigenbaum's requirements. This statistical based methodology can also be used to derive vectors for the neural network inputs.

The method discussed in this paper is implemented using the initial phonemic knowledge model. The implementation includes the refining of the phonetical data to a map that is derived from the clusters. The map consists of knowledge units, and the connection between units provides information on the conditional probability of a unit in the context of its predecessors.

Figure 5 shows an implementation of phonemic knowledge classification. In the example given the group C has the class k as one of its members. The most likely phoneme found in the given contexts to follow a class k of group C is subclass D'. Subclasses m, n and s are most likely to follow subclass D' in this sequence.

The method of phonemic knowledge introduced here has contributed new insights into how pattern would recognised. It appears to have the potential to address the following two issues. The first is how to acquire and refine pattern knowledge. The second is how to equip a system with the capability to adapt and evolve. The phonemic knowledge is regarded as bottleneck in building an intelligent recognisor. The model discussed in this paper introduces a means of utilising this knowledge to aid the recognition process.

Other hierarchy clusters still under study. Those are derived from the statistical characteristics of phonemes signals which would reflected on the neural network architecture, and produce another knowledge representation structure in the system as pattern knowledge, where the inputs provided to the network are derived from the statistical characteristics of speech signal.

REFERENCES:

Fiegenbaum E. A., Cohen P.R. "The Handbook of Artificial Intelligence". Vol 1-3. Addison Wesley publishing company. 2ed. edition. 1989.

Hanazawa T., Hinton G., Shikano K. "Phoneme Recognition Using Time Delay Neural Networks". IEEE transaction on Acoustic Speech and Signal Processing. March 1989.

Katsuhiko Shirai, Naoki Hosaka, and Eichiro Kitagowa. "Speaker Adaptive Phoneme Recognition". Proceeding ICASSP-90. p-p 169-72.

McDermott E., Katagiri S. "Shift-Invariant Multicategories Phoneme Recognition Using Cohonen's lvq2". IEEE International conference on Acoustic Speech and Signal Processing, May 1989.

Morris W. Firebaugh (editor). " The Artificial Intelligence -A Knowledge Based Approach". PWS Kent publishing company. Boston U.S.A. Chapter 9. 1989.

Sawai H., Shikano K. "Integrated Training For Spotting Japanese Phoneme Using Large Phonemic Time-Delay Neural Network". IEEE International Conference on Acoustic, Speech and Signal Processing. May 1990.

Tam K. Y. "Applying Conceptual Clustering to Knowledge Basis Construction". The international Journal of Decision Support Systems. Vol 10. No 2. September 1993.

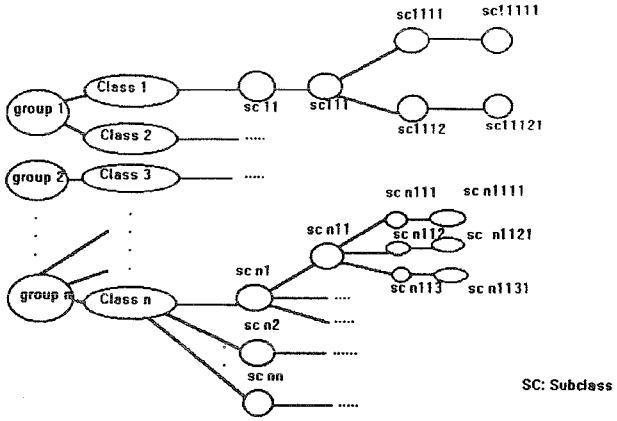

Fig 1 The Hierarchical structure of the phonemic kowledge.

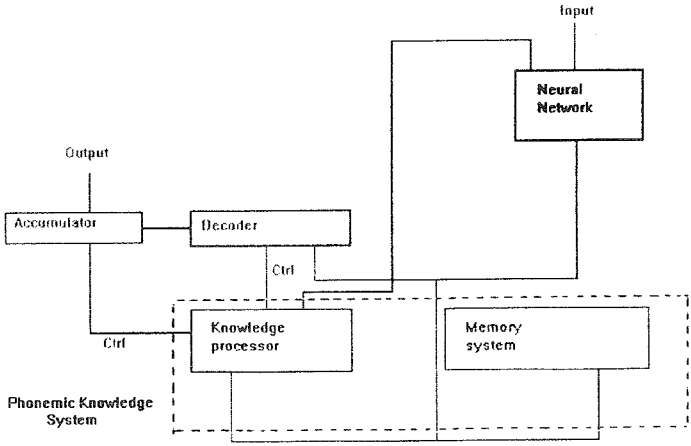

Figure 2- Block diagram of the speech recognition system

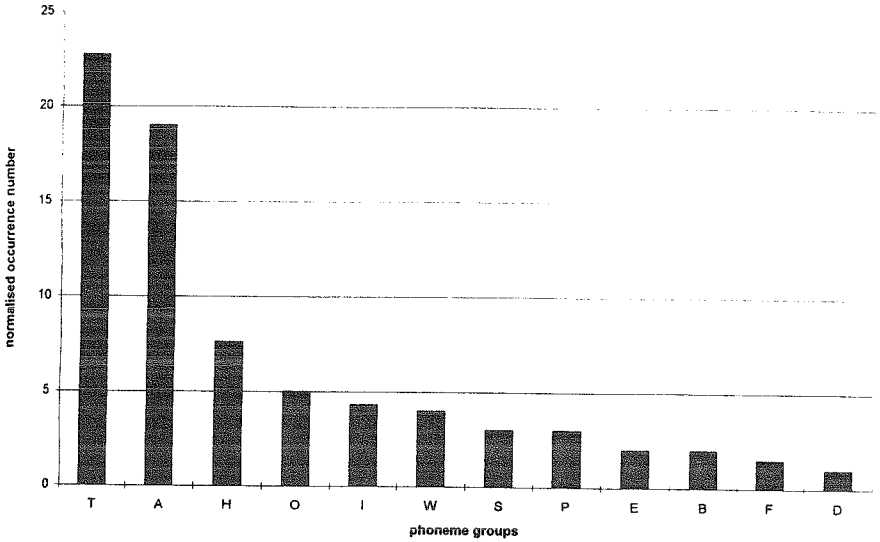

Fig.3. The likelihood of phoneme groups occurrence for the starting position in the first stage of the knowledge

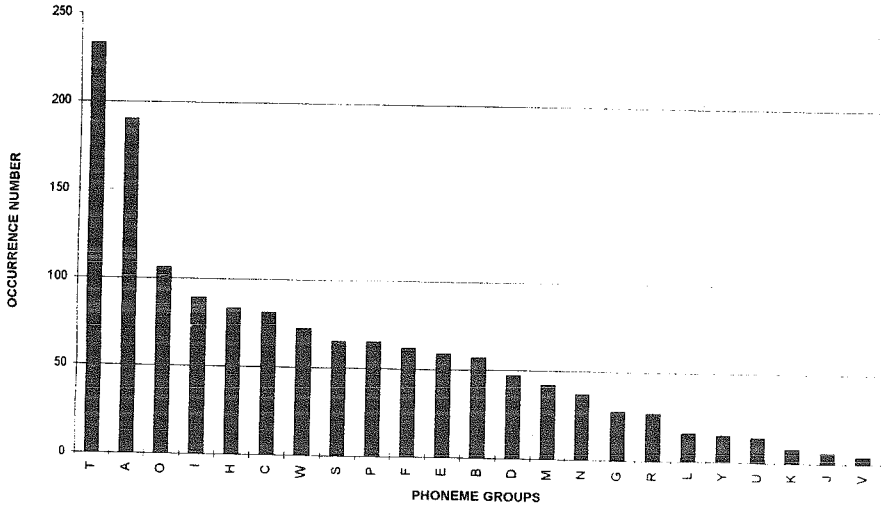

Fig.4. The likelihood of phoneme groups occurrence for the strarting position in the second stage of the knowledge