

ABSTRACT

The development of LVCSR needs a sentence databases which is rich of voices. In this case, voices are represented as phoneme. Phoneme is the smallest voice component in spelling the word that it will be arranged to be threephones. The aim is to increase the accuracy of conversion of speech to text. The system of this research paper discusses about selecting sentences which has rich of threephones.

Greedy algorithm is an algorithm selects current best solution. It is hoped that it will gets the best solution on the last execution, although there's no guaranty that it will always get the best solution. This algorithm will be used to select some sentences from the whole sentences set. The sentences selected is the triphone rich sentences. The selection process needs big memory allocation. So how we can do the selection with a minimal memory allocation is to do the process to the sub collection of the whole sentences set, in other words, we will break apart the whole sentence set to some sub sentence set and process it separately.

Keywords: *databases. Phoneme. Threephones. Conversation.*