

# **A Study in Speaker Dependent Medium Vocabulary Word Recognition: Application to Human/Computer Interface**

By

Moatassem M. Abdallah Abd El Salam

## **(ABSTRACT)**

**H**uman interfaces to computers continue to be an active area of research. The keyboard is considered the basic interface for editing control as well as text input. Problems of correct typing and typing speed have urged research for alternative means for keyboard replacement, or at least “resizing” its monopoly. Pointing devices (e.g. a mouse) have been developed, and supporting software with icons is now widely used. Two other means are being developed and operationally tested, namely, the pen for handwriting text, commands and drawings, and spoken language interface, which is the subject of this thesis.

Human/computer interface is an interactive man-machine communication facility that enjoys the following advantages

- High input speed: some experiments reveal that the rate of information input by speech is three times faster than keyboard input and eight times faster than inputting characters by hand.
- No training needed: because the generation of speech is a very natural human action, it requires no special training.
- Parallel processing with other information: production of speech works quite well in conjunction with gestures of hands and feet for visual perception of information.
- Simple and economical input sensor: microphones are inexpensive and are readily available.

- Coping with handicaps: these interfaces can be used in unusual circumstances of darkness, blindness, or other visual handicap.

This dissertation presents a design of a Human Computer Interface (HCI) system that can be trained to work with an individual speaker. A new approach is introduced to extract key voice features, called Median Linear Predictive Coding (MLPC). MLPC reduces the HCI calculation time and gives an improved recognition rate. This design eliminated the typical Multi-layer Perceptron (MLP) problems of complexity growth with vocabulary size, the large training times required and the need for complete re-training whenever the vocabulary is extended. A novel modular neural network architecture, called a Pyramidal Modular Neural Network (PMNN), is introduced for recursive speech identification. In addition, many other system algorithms/components, such as speech endpoint detection, automatic noise thresholding, etc., must be tailored correctly in order to achieve high recognition accuracy.