INTRODUCTION TO SPEECH RECOGNITION

INTRODUCTION

Speech recognition is an interdisciplinary subject. In the past two decades, speech recognition technology has made significant progress and has begun to move from the laboratory to the market. The fields involved in speech recognition technology include: signal processing, pattern recognition, probability theory and information theory, vocal mechanism and auditory mechanism, artificial intelligence and so on.

I. IDENTIFY PATTERNS

Soviet research laid the foundation for the application of pattern recognition to the field of speech recognition;Japanese research shows how to use dynamic programming techniques to perform nonlinear temporal matching between to-be-recognized speech patterns and standard speech patterns; Itakura's research proposes how to extend Linear;Predictive Analysis (LPC) to feature extraction of speech signals.

II. DATABASE

In the process of research and development of speech recognition, relevant researchers have designed and produced speech databases in Chinese (including different dialects), English and other languages according to the pronunciation characteristics of different languages. These speech databases can be used for domestic and foreign scientific research. Units and universities provide sufficient and scientific training speech samples for Chinese continuous speech recognition algorithm research, system design, and industrialization. For example: MIT Media lab Speech Dataset (Massachusetts Institute of Technology Media Lab Speech Dataset), Pitch and Voicing Estimates for Aurora 2 (Gene cycle and pitch estimation for Aurora2 speech library), Congressional speech data (Congressional speech data), Mandarin Speech Frame Data (Mandarin speech frame data), speech data for testing blind source separation algorithms, etc.

III. IDENTIFICATION METHOD

Speech recognition methods are mainly pattern matching methods. In the training phase, the user speaks each word in the vocabulary in turn, and stores its feature vector as a template in the template library. In the recognition stage, the similarity between the feature vector of the input speech and each template in the template library is compared in turn, and the one with the highest similarity is output as the recognition result.

IV. ACUSTIC SIGNATURE

The extraction and selection of acoustic features is an important part of speech recognition. The extraction of acoustic features is not only a process of greatly compressing information, but also a process of signal deconvolution, in order to make the mode divider better. Due to the time-varying nature of speech signals, feature extraction must be performed on a small segment of speech signals, that is, short-term analysis. This period of analysis is considered to be stable and is called a frame, and the offset between frames is usually 1/2 or 1/3 of the frame length. The signal is usually pre-emphasized to boost high frequencies, and the signal is windowed to avoid the effects of short-term speech segment edges.

V. LANGUSGE MODEL

The performance of language models is usually measured by cross-entropy and complexity (Perplexity). The meaning of cross-entropy is the difficulty of text recognition with the model, or from a compression point of view, how many bits are used to encode each word on average. The meaning of complexity is to use the model to represent the average number of branches of this text, and its reciprocal can be regarded as the average probability of each word.

SUMMARY

The foregoing describes techniques for implementing various aspects of a speech recognition system. These techniques have achieved good results in practical use, but how to overcome various factors that affect speech still needs to be analyzed more deeply. At present, the dictation machine system cannot be fully practical to replace the keyboard input, but the maturity of recognition technology has also promoted the research of higher-level speech understanding technology.

Zhong Wu, undergraduate's student in the Faculty of Information Technology and Management of BSUIR, 2921123673@qq.com.

Jiang Shuqin, undergraduate's student in the Faculty of Information Technology and Management of BSUIR, chenphon@163.com.

Trofimovich A.F., Senior Lecturer of Information Technologies in Automated Systems Department, trofimaf@bsuir.by