ISSN:2320-0790

An International Journal of Advanced Computer Technology

A process to improve the accuracy of voice recognition system by using word correction system

Mr. Nazrul Islam[1], Mr. Kamal Kumar Ranga[2]

[1]Computer Science & Engineering Department, Ganga Institute of Technology & Management, India.
[2]Computer Science & Engineering Department, Ganga Institute of Technology & Management, India.

Abstract: Recently, computer speech recognition is used to solve problems and any plan based task, predictable features of the user's behaviour may be inferred and used to aid the recognition of the speech input. The MINDS system generates expectations of what will be said next and uses them to assist speech recognition. Since a user does not always conform to system expectations, MINDS handles violated expectations. We use a common knowledge to enable the speech system to give priority to recognizing what a user is most likely to say. Each time the words spoken by speaker will display in the computer which is generated by the speech recognizer. In it words correction system has been implemented. With the help of words correction system, speaker can correct the word manually through keyboard, if the word produced by the voice recogniser is not similar as the speaker dictates to the computer. The word correction system is used to correct misrecognised words. This system will bring cent percent accuracy in voice recognition system.

Keywords: Algorithm, Speech, windows speech recognition devices.

I. INTRODUCTION

Speech recognition, the ability to identify spoken words, and speaker recognition, the ability to identify who is saying them, become commonplace applications of speech processing technology. Limited forms of speech recognition are available on personal workstations. Recently there is much interest in speech recognition. The performance is improving and Speech recognition has already became useful for many applications, such as telephone voice-response systems for selecting services or information, digit recognition for cellular phones, and data entry while walking around a railway yard or clambering over a jet engine during an inspection. Nonetheless, comfortable and natural communication in a general setting (no constraints on what you can say and how you say it) is beyond us for now, posing a problem too difficult to solve [1]. Fortunately, we can simplify the problem to allow the creation of applications for word correction system. Voice recognition is related to work on speech recognition. Instead of determining who said it, you determine what was said. The work of determining what was said by a speaker is called word recognition. The most general form of word recognition is still not very accurate for all voice recognition systems.

populations, but if you constrain the words spoken by the user (text-dependent) and do not allow the speech quality to vary too wildly, then it too can be done on a workstation Early attempts to design systems for automatic speech recognition were mostly guided by the theory of acoustic-phonetics, which tells the elements of speech and how they are realized to form a spoken language. In 1952, Davis et al. of Bell Laboratories built a model for isolated digit recognition for a single speaker, using the spectral resonances during vowel regions of each digit. In 1956, Olson and Belar of RCA Laboratories tried to recognize ten syllables of a single speaker. At MIT Lincoln Laboratory, Forge and Forge built a speaker-independent ten-vowel recognizer in 1959. They used time-varying estimates of the vocal tract resonance. Later, in the 1960s, with emphasis on building a special hardware, several Japanese laboratories also demonstrated their progress. Most notable among them were the vowel recognizer of Suzuki and Nakata of the Radio Research Lab in Tokyo, the phoneme recognizer of Sakai and Doshita of Kyoto University (noting the use of a speech segmenter to allow analysis and recognition of speech in different portions of the signal), and the digit recognizer of NEC Laboratories. One significant remark to be made is the year 1959 when Fry and Denes, at University College in England, attempted a phoneme recognizer to recognize four vowels and nine consonants. They incorporated statistical information about allowable phoneme sequences in English to enhance the overall phoneme recognition accuracy for words consisting of two or more phonemes [2]. This may marked the first use of statistical syntax in automatic speech recognition. The work of Martin’s team at RCA Laboratories and that of Vintsyuk in the Soviet Union in the 1960s have particularly important implications on the research and development of automatic speech recognition. Martin recognized the need to deal with the non uniformity of time-scale in speech events and suggested realistic solutions, including detection of utterance endpoints, which greatly enhanced the reliability of the recognizer performance. Vintsyuk proposed the use of dynamic programming for time-alignment between two utterances in order to derive a meaningful matching score. Although his work was largely unknown to the West then, it appears to have preceded that of Sakoe and Chiba, as well as others who proposed more formal methods in speech pattern matching, generally known as dynamic time warping. Since the late 1970s, dynamic programming, in numerous variant forms, has become an indispensable technique in the pattern-matching approach to automatic speech recognition.

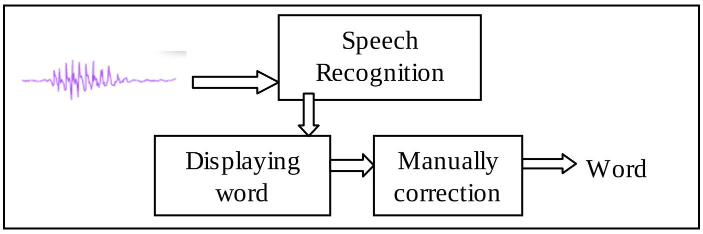

Fig. 1. Extracting information from speech to find accurate word

II. HISTORY & UNDERSTANDING OF SPEECH RECOGNITION

The Spoken language processing encompasses a broad range of technical challenges, including recognition of words and phrases in the speech signal, extraction of keywords or key phrases in the utterance, and understanding of the spoken utterance for the machine to take actions. Conversation between people can take many different forms, many of which may be beyond the scope of the current scientific interest. For example, a casual conversation between two people can drift over an unbounded domain with no end result anticipated. We will not address this category of scenarios [4]. We will, however, assume that the common goal in speech recognition and understanding is to identify an important message, out of a finite set of possibilities, conveyed in the spoken utterance.

A. *Fundamentals of Linguistics and Acoustic-Phonetics*

Your Most of the classical speech-recognition research was based on the identification paradigm as discussed above. It requires extensive understanding of the properties of the object (i.e., the speech sound). It, thus, depends on and makes use of, almost exclusively, the acoustic-phonetic theory, which aims at building a framework for understanding speech by a human [5].

Phoneticians and linguists decompose a spoken language into elements of linguistically distinctive sounds—the phonemes. The number of phonemes in a language is often a matter of judgment and is not invariant to different linguists. Phonemes are determined and taxonomically classified according to their corresponding articulator configurations. For example, a vowel is produced by exciting a vocal tract of an essentially fixed shape with quasi-periodic pulses of air, caused by the vibration of the vocal cords. Front vowels (/i/, /I/, /e/, and /E/) are vowels produced with a tongue hump in the front portion of the vocal tract. Other phoneme categories include diphthongs, semivowels, nasals, stops, fricatives, affricates, and whisper. As in many classical studies, the taxonomy was established for a systematic investigation of the properties of the "element" of speech sounds. Such properties of sounds are often referred to as acoustic-phonetic features [6]. An alternative way to classify the phonemes is to use the broad phonetic class according to key acoustic-phonetic feature dimensions.

B. *Factors Affecting Speech Recognition*

Modern speech recognition research began in the late 1950s with the advent of the digital computer. Combined with tools to capture and analyze speech, such as analog-to-digital converters and sound spectrograms, the computer allowed researchers to search for ways to extract features from speech that allow discrimination between different words. The 1960s saw advances in the automatic segmentation of speech into units of linguistic relevance (such as phonemes, syllables, and words) and on new pattern-matching classification algorithms. By the 1970s, a number of important techniques essential to today's state-of-the-art speech recognition systems had emerged, spurred on in part by the Defence Advanced Research Projects Agency speech recognition project. These techniques have now been refined to the point where very high recognition rates are possible, and commercial systems are available at reasonable prices.

C. *Statistical Pattern Recognition Formulation- Data-Driven Approach*

The formulation of statistical pattern recognition has its root in Bayes' decision theory [8]. Let X be a random observation from an information source, consisting of M classes of event. A classifier's job is to correctly classify each X into one of the M classes. (Here, we use the terms classifier and recognizer interchangeably because we have defined the problem as identifying an unknown observation as one of M classes of event.) We denote these classes by Ci, i=1,2,…,M. Let P(X,Ci) be the joint probability distribution of X and Ci, a quantity that is assumed to be known to the designer of the classifier. In other words, the designer has full knowledge of the random nature of the source.

$$R(C_i|\mathbf{X}) = \sum_{j=1}^{M} e_{ij} P(C_j|\mathbf{X}) \mathbf{1}(\mathbf{X} \in C_j) \qquad (1)$$

To measure the performance of the classifier, we further define for every class pair (i, j) a cost or loss function eij , which signifies the cost of classifying (or recognizing) a class i observation into a class j event [9]. The loss function is generally nonnegative, with eij=0 representing a correct classification.

III. TECHNOLOGY COMPONENTS OF NORMAL SPEECH RECOGNITION & UNDERSTANDING

Before any machine can interpret speech, a microphone must translate the vibrations of a person's voice into a wavelike electrical signal. This signal in turn is converted by the system's hardware for instance, a computer's sound card, into a digital signal. It is the digital signal that a speech recognition program analyzes in order to recognize separate phonemes, the basic building blocks of speech. The phonemes are then recombined into words. However, many words sound alike, and, in order to select the appropriate word, the program must rely on the context. All Most computer systems for speech recognition include the following five components.

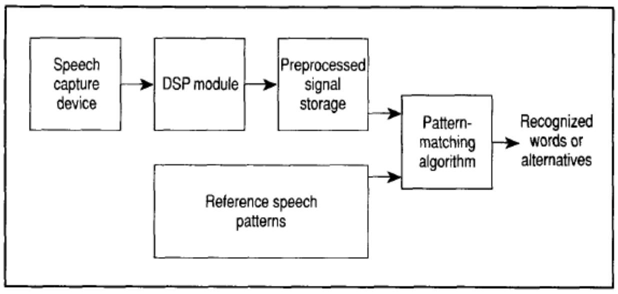

Fig. 2. Components of a normal speech recognition system

This usually consists of a microphone and associated analog-to-digital converter, which digitally encodes the raw speech wave form.

The DSP module performs endpoint (word boundary) detection to separate speech from non speech, converts the raw waveform into a frequency domain representation, and performs further windowing, scaling, filtering, and data compression. The goal is to enhance and retain only those components of the spectral representation that are useful for recognition purposes, thereby reducing the amount of information that the pattern-matching algorithm must contend with. A set of these speech parameters for one interval of time (usually 10-30 milliseconds) is called a speech frame.

Here, the preprocessed speech is buffered for the recognition algorithm. Stored reference patterns can be matched against the user's speech sample once it has been preprocessed by the DSP module. This information is stored as a set of speech templates or as generative speech models. The algorithm must compute a measure of goodness-of-fit between the preprocessed signal from the user's speech and all the stored templates or speech models. A selection process chooses the template or model (possibly more than one) with the best match.

IV. PERSONALIZING THR WORD CORRECTION SYSTEM

In this paper, our concept will help to increase the accuracy of the voice recognition System. In the words correction system speaker can correct the word manually through keyboard, if the word produced by the voice recogniser is not similar as the speaker dictates to the computer. The word correction system is used to correct misrecognised words. The accurate correction of misrecognised words is the key to improve the user's voice-recognition experience. Here, simply editing misrecognised text will improve recognition accuracy.

V. CONCLUSION

We are using word correction system to find the word misrecognized by the voice recognition system and it will increase the accuracy of voice recognition system. Besides, we are giving option to correct misrecognised words produced by the speech recognizer. For example, if a word 'misguide' is initially transcribed as 'Miss guide', the word 'Miss guide'can be corrected by the speaker.

VI. ACKNOWLEDGMENT

We wish to thanks K. Lee- the member of the IEEE, H. Hon and R. Reddy- fellow of the IEEE. We would like to acknowledge L.R. Bahl. We would also like to thank R. Bakis, P.S. Cohen, A. Cole, Jelinek, B.L. Lewis, and R.L. Mercer of IBM T.J. Watson Research Center, Yorktown Heights, NY 10598.

VII. REFERENCES

[1] K. Lee, Automatic, Speech Recognition: the development to f the Sphinx System, Kluwer Academic Publishers, Norwell, Mass. 1989

[2] L. R. Bahl et al., "Speech Recognition of a Natural Text Read as Isolated Words," Proc. IEEE IntI Conf. Acoustics, Speech, and Signal Processing, April 1981, pp. 1,168-1,17 1.

[3] D.O. Kimbal et al., "Recognition Performance and Grammatical Constraints," Proc. DARPA Speech Recognition Workshop, Feb. 1986, pp. 53-59.

[4] D. O'Shaughnessy, Speech Communication: Human and Machine, Addison- Wesley, Reading, Mass., 1987.

[5] M.A. Franzini, M.J. Witbrock, and K.-F. Lee, "Speaker-Independent Recognition of Connected Utterances Using Recurrent and Non recurrent Neural Networks," Proc. Int'l Joint Conf. Neural Networks, V01.2, Washington, DC, June 1989, pp.11-1 to II-

[6] J. Mariani, "Recent Advances in Speech Processing," Proc. IEEE Intl Conf. Acoustics, Speech, and Signal Processing, Glasgow, Scotland. May 1989, pp. 429- 440.

[7] M.-W. Fung et al., "Improved Speaker Adaptation Using Text-Dependent Spectral Mappings," Proc. IEEE Int'l Conf. Acoustics, Speech. and Signal Processing, New York City, 1988, pp. 131-134.

[8] D.B. Paul, "The Lincoln Robust Continuous Speech Recognizer," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Glasgow, Scotland, 1989, pp. 449-452.

[9] H. Murveit and M. Weintraub, "1,000- Word Speaker-Independent Continuous- Speech Recognition Using Hidden Markov Models," Proc. IEEE Int'l Conf, Acoustics, Speech, and Signal Processing, New York City, 1988, pp. 115-1 18.

[10] W. Wylegala, "A 20,000-Word Recognizer Based on Statistical Evaluation Methods," Speech Technology Msagazine, Apr./May 1989, pp, 16-18.