Trends in

Applied Sciences Research

ISSN 1819-3579

Academic Journals Inc.

www.academicjournals.com

Trends in Applied Sciences Research 1 (1): 79-82, 2006
ISSN 1819-3579

Voice Identification Based on HMMs

Li Jing, Huang Hua and Liu Sufang
School of Electrical Engineering and Information Technology,
Sichuan University Chengdu 610065, China

Abstract: This study discusses the definition of voice identification and describes how the speaker recognition system makes up of. At present, there are five specific forms of voice identification technology, as for the methods of voice identification, the most essential one is Hidden Markov Models (HMM) which is relatively effective. One more important way to describe HMM is shown simply. And then introduces applications of voice identification, in which an innovative one is voice security system.

Key words: Biometric identification, voice identification technology, HMM, voice security, false reject rate

Introduction

Along with the development of increasing electronic equipments, security becomes the most important problem of many areas. Because of the uniqueness, biometric technologies, such as voice, fingerprints, are defined as 'automated methods of identifying or authenticating the identity of a living person'. As one aspect of biometric identification, there are many advantages to using voice identification: reliability, flexibility, eliminating spelling errors and providing eyes and hands-free operation. Its convenience and easiness to accept will make it more extensive in security system, telephone application and more useful to people with disabilities (McDermott *et al*., 1996). However, the difficulty that noisy voice is existent is obvious (Raj *et al*., 2001). This study concentrates on the essential method-HMM, which can realize voice identification effectively. Voice identification involves enrollment and verification. At enrollment, the person offers a voice sample. This is canned electronically, processed and stored as a template.

Voice Identification System

Voice identification which can predict whether the person is right is essentially pattern identification. The voice of everyone will be influenced by individual build, dialect and environment. Although it is advanced, there are also some disadvantages like other behavioral traits. For instance, the change scope is so broad that it is difficult to carry out accurate matching; the voice will affect the result of collection and compare if the volume, rate and timbre are changing.

Figure 1 shows the signal processing of voice identification system (Picone, 1996). A voice identification system mainly has three parts: segmentation of voice signal, pickup of characteristics and voice identification. Voice identification model can be divided into parametric model usually using HMM (Hidden Markov Model) and nonparametric model often using ANN (Artificial Nerve Net).

Corresponding Author: Li Jing, School of Electrical Engineering and Information Technology, Sichuan University Chengdu 610065, China Tel: 86 028-85466088

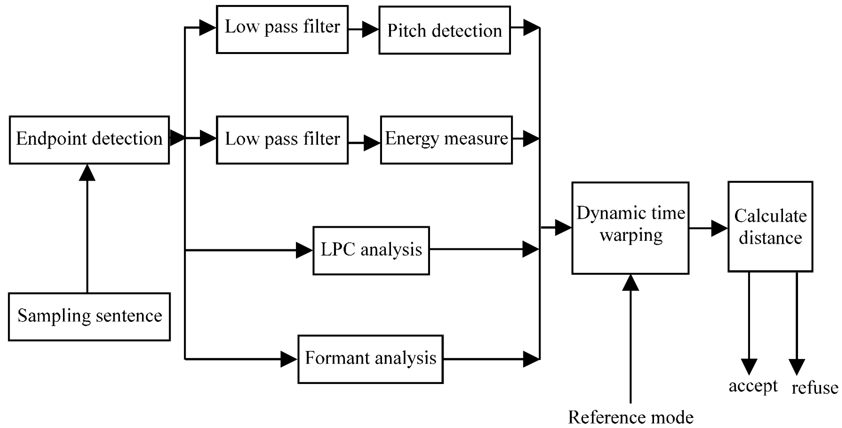

Fig. 1: Signal processing of voice identification

HMM could describe dynamic time sequence preferably; static classify ability of ANN is better. However, the two models are not independent drastically, which can also be combined.

Method of Voice Identification

There are five specific forms of voice identification technologies that are currently available or under development.

The first is speaker dependent; this type of technology involves 'training' the system to recognize your speech patterns. It is the best if used by a specific user. The second is speaker independent; this type of voice identification technology can be used by anyone without having to train the system. The third is discrete speech input; this environment involves the person speaking to make small pauses, as small as 1/10 of a second, between words. The forth is continuous speech input, users can speak at a continuous rate but the voice identification software can only recognize a limited amount of words and phrases. The last is natural speech input, this is the most desired form of voice identification, but is still under development. Here the user is able to speak freely and the system is able to interpret and carry out commands on-the-fly. Aiming at foregoing forms of recognition technology, we bring forward an essential method-HMM.

Hidden Markov Models (HMMs) are a popular approach for speech recognition. HMMs are commonly used in some other applications, such as handwriting recognition, gene sequencing and even rainfall. In the speech community, HMMs are used to represent the joint probability distribution of a collection of speech feature vectors.

There is an important property about an HMM, each new sample of observations requires a completely new sample from the hidden Markov chain, namely, two different observation samples from an HMM will typically comes from two different underlying state assignments to the hidden chain.

Generally, there are three methods to graphically describe HMMs (Bilmes, 1999). Figure 2 shows the main way that only be discussed in this paper. We can see that the Markov-chain topology is not specified --only the HMM conditional independence properties are shown. Along with the introducing of random variables, this view of an HMM is preferable when discussing the statistical dependencies directly represented by an HMM.

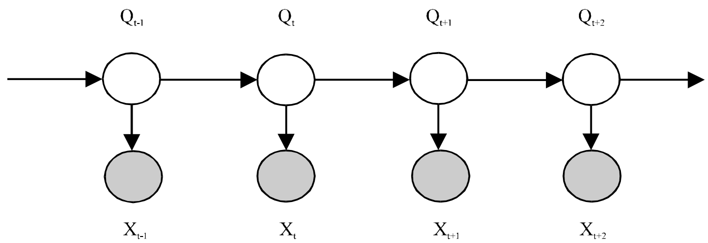

Fig. 2: The Way to Describe an HMM
$Q_{1:T}$ is the "hidden" variables which are discrete
$X_{1:T}$ is the "observed" variables which may be either discrete or continuous

An HMM is expressed by a parameter group $\lambda = \{\pi, A, B\}$, which is used to describe the probability model of statistic characteristic of a random sequence. π is initial distribution to describe the state q of observation sequence at time t, $\pi_1 = P(q_1 = S)$, $i = 1,2,\Lambda, N$, it is the probability distribution of each state. $A= \{a_{ij}| i, j = 1, 2, \Lambda, n\}$ is the homogeneous stochastic matrix representing the transition probabilities, $a_{ij} = P (q_1 = j \mid q_{t-1} = i)$ shows that the sample at time t is obtained from the state assignment at time t-1, is conditionally independent of the previous hidden and observation variables. B is the collection of observation probability distributions, $bj(x) = p (Xt = x|Qt = j)$. (Suifang *et al*., 2005).

The samples from an HMM could be obtained as: $q_t = i$ with prob. $p(Q_t = i \mid q_{t-1})$ and $x_t \sim b_{qt} (x)$ It means that only the hidden state assignment at time t is used to determine the observation distribution at that time. In HMM, a new sample of observations requires a completely new sample from the hidden Markov chain, this is an important fact of HMM.

Applications

Voice identification based on correlative science principles is friendly. Speech identification is sometimes helpful for environments that are hands-busy, eyes-busy, mobility-required, or hostile and shows promise for telephone-based services.

Most applications of voice identification has penetrated recently include medium-security access control and time and attendance monitoring. Voice security systems are responsible for an innovative method of security that dramatically reduces fraud and can insure ones property from use, even if stolen, or obtained fraudulently. Since September 11th 2001, the world changed forever and we clearly know that security is the number one priority in our lives. Fraud is the USA's number one problem and rapidly spreading throughout the world. Examples of usage of this technology are cell phones (to eliminate cell phone fraud), ATM manufacturers and automobile manufacturers (to dramatically reduce theft and car jacking).

The following applications are the most common right now (http://et.wcu.edu/aidc/BioWebPages/Biometrics_Voice.html).

- General Motors uses voice identification systems to restrict access to some of its computer room.
- Martin Marietta, GM and Hertz are using voice identification technology to protect their computer facilities.
- Private estates all over the world are protected with voice identification technology.

- Used in telephone security-based applications.
- Staff at hospitals needs to pass a voice system to enter some important units.
- Immigration and Naturalization Service has implemented voice identification for frequent travelers that cross border.
- Many people have difficulty typing due to physical limitations such as repetitive strain injuries (RSI), muscular dystrophy and many others. For example, people with difficulty hearing could use a system connected to their telephone to convert the caller's speech to text.

Discussion

In the 1960's, the basis for voice identification technology was pioneered by Texas Instruments. Since that time, voice identification has undergone aggressive research and development to bring it into mainstream society. Today voice identification analysis has matured into a sophisticated identification technology. HMM in this study is an essential technology by which some new methods are developed. Error rates that use this type of language modeling are from one to 15% (Richardson *et al*., 2003). However, voice identification technology is still slow to evolve in many areas. One main reason is that it is not as accurate as other biometric technologies. Mostly because of background noise, they tend to have a high false reject rate and its accuracy is not as better as iris. Although voice identification is not the most accurate, its 'natural' and the more mature theory make for wider applications. Nowadays, voice identification is developing from continuous voice input to natural voice input and has realized 97% of accuracy of continuous voice identification. It has been estimated that if voice identification technology continues to progress as it has, keyboards will become obsolete in ten years.

References

Bilmes, J.A., 1999. Natural statistical models for automatic speech recognition. http://www.icsi.berkeley.edu/~bilmes/papers/tr-99-016.pdf.

http://et.wcu.edu/aidc/BioWebPages/Biometrics_Voice.html.

McDermott, M.C., T. Owen and F.M. McDermott, 1996. Voice Identification: The Aural/Spectrographic Method. 1996. http://www.owlinvestigations.com/forensic_articles/aural_spectrographic/fulltext.html

Picone, J., 1996. Fundamentals of Speech Recognition: A Short Course. Texas Instruments. May 15-17, 1996.

Raj, B., M.L. Seltzer and R.M. Stern, 2001. Robust Speech Recognition: The Case for Restoring Missing Features. http://research.microsoft.com/~mseltzer/papers/mseltzer_crac01_missfeat.pdf

Richardson, M., J. Bilmes and C. Diorio, 2003. Hidden-articulator markov models for speech recognition. Speech Communication, 41: 511-529.

Suifang, L. and P. Yongxiang *et al*., 2005. Noisy speech recognition based on hybrid model of hidden markov models and wavelet neural network. J. System Simulation, July 2005.