

Chapter 13

Speaker Verification and Identification

Minho Jin

Korea Advanced Institute of Science and Technology, Republic of Korea

Chang D. Yoo

Korea Advanced Institute of Science and Technology, Republic of Korea

ABSTRACT

A speaker recognition system verifies or identifies a speaker's identity based on his/her voice. It is considered as one of the most convenient biometric characteristic for human machine communication. This chapter introduces several speaker recognition systems and examines their performances under various conditions. Speaker recognition can be classified into either speaker verification or speaker identification. Speaker verification aims to verify whether an input speech corresponds to a claimed identity, and speaker identification aims to identify an input speech by selecting one model from a set of enrolled speaker models. Both the speaker verification and identification system consist of three essential elements: feature extraction, speaker modeling, and matching. The feature extraction pertains to extracting essential features from an input speech for speaker recognition. The speaker modeling pertains to probabilistically modeling the feature of the enrolled speakers. The matching pertains to matching the input feature to various speaker models. Speaker modeling techniques including Gaussian mixture model (GMM), hidden Markov model (HMM), and phone n-grams are presented, and in this chapter, their performances are compared under various tasks. Several verification and identification experimental results presented in this chapter indicate that speaker recognition performances are highly dependent on the acoustical environment. A comparative study between human listeners and an automatic speaker verification system is presented, and it indicates that an automatic speaker verification system can outperform human listeners. The applications of speaker recognition are summarized, and finally various obstacles that must be overcome are discussed.

DOI: 10.4018/978-1-60566-725-6.ch013

INTRODUCTION

Speaker recognition can be classified into either 1) speaker verification or 2) speaker identification (Furui, 1997; J. Campbell, 1997; Bimbot et al., 2004). Speaker verification aims to verify whether an input speech corresponds to the claimed identity. Speaker identification aims to identify an input speech by selecting one model from a set of enrolled speaker models: in some cases, speaker verification will follow speaker identification in order to validate the identification result (Park & Hazen, 2002). Speaker verification is one case of biometric authentication, where users provide their biometric characteristics as passwords. Biometric characteristics can be obtained from deoxyribonucleic acid (DNA), face shape, ear shape, fingerprint, gait pattern, hand-vein pattern, hand-and-finger geometry, iris scan, retinal scan, signature, voice, etc. These are often compared under the following criteria (Jain, Ross, & Prabhakar, 2004):

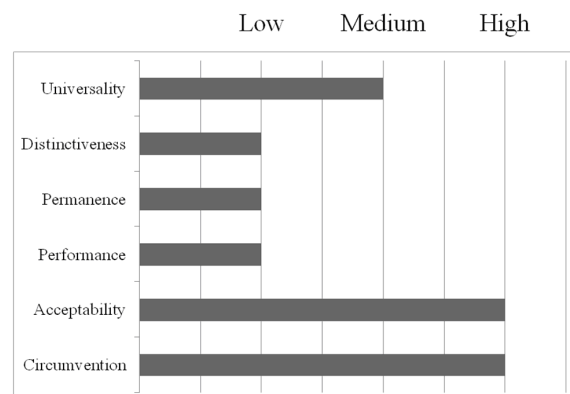
- **Universality:** the biometric characteristic should be universally available to everyone.
- **Distinctiveness:** the biometric characteristics of different people should be distinctive.
- **Permanence:** the biometric characteristic should be invariant over a period of time that depends on the applications
- **Performance:** the biometric authentication system based on the biometric characteristic should be accurate, and its computational cost should be small.
- **Acceptability:** the result of a biometric authentication system based on certain biometric characteristic should be accepted to all users.

One additional criterion that should be included is circumvention which is given by

- **Circumvention:** biometric characteristics that are vulnerable to malicious attacks are leading to low circumvention.

High biometric characteristic scores on all above criteria except circumvention are preferable in real applications. As shown in Figure 1, voice is reported to have medium universality. However, in many cases, voice is the only biometric characteristic available: for example, when a person is talking over the phone. The distinctiveness of voice is considered low, and very often a speaker verification system can be fooled by an impostor mimicking the voice of an enrolled. For this, many features such as prosodic and idiosyncratic features have been incorporated to improve the speaker recognition system. The permanence of voice is low since a speaker's voice can vary under various situations, physical conditions, etc. By incorporating on-line speaker adaptation techniques that adapt a speaker's voice change on-line, the permanence of voice can be improved. We discuss the performance of a speaker recognition system in the latter part of this chapter.

Figure 1. Properties of voice: voice can be universally available for every person, and its authentication result is acceptable. However, its performance in terms of accuracy is known to be slightly inferior to that of other biometric characteristics



24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/speaker-verification-identification/36924

Related Content

Brain-Computer Interfaces and Visual Activity

Carmen Vidaurre, Andrea Kübler, Michael Tangermann, Klaus-Robert Müller and José del R. Millán (2012). *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (pp. 153-174).

www.irma-international.org/chapter/brain-computer-interfaces-visual-activity/60039

Human Face Region Detection Driving Activity Recognition in Video

Anastasios Doulamis, Athanasios Voulodimos and Theodora Varvarigou (2014). *Face Recognition in Adverse Conditions* (pp. 316-335).

www.irma-international.org/chapter/human-face-region-detection-driving-activity-recognition-in-video/106989

A Study of Sparse Matrix Methods on New Hardware: Advances and Challenges

Athanasios Fevgas, Konstantis Daloukas, Panagiota Tsompanopoulou and Panayiotis Bozanis (2015). *International Journal of Monitoring and Surveillance Technologies Research* (pp. 1-19).

www.irma-international.org/article/a-study-of-sparse-matrix-methods-on-new-hardware/146152

Facial Expression Recognition

Daijin Kim and Jaewon Sung (2009). *Automated Face Analysis: Emerging Technologies and Research* (pp. 255-317).

www.irma-international.org/chapter/facial-expression-recognition/5475

Rank Level Fusion

(2013). *Multimodal Biometrics and Intelligent Image Processing for Security Systems* (pp. 69-79).

www.irma-international.org/chapter/rank-level-fusion/76162