
Prolog to

Speaker Recognition: A Tutorial

An introduction to the paper by Campbell

Two classes of applications are developed based on speaker-recognition algorithms: speaker identification and speaker verification. Speaker-identification applications attempt to answer the question, “Who are you?” Speaker-verification applications attempt to answer the question, “Are you whom you claim to be?” Various automated biometric-based identification and verification systems have been developed to answer these questions, some of which perform quite well in terms of accuracy and speed. One might wonder, “Why do we need one more biometric?”

In contrast to other automated biometrics-based systems, which are mostly image based, a speaker identification or verification system is based on voice and spoken words. This distinction affords two valuable characteristics: non-invasiveness and readily available data acquisition and transmission devices (e.g., telephone). As a biometric verification application user, one normally does not object to speaking a few words or phrases to gain access to a facility, network, or data base. For example, a credit card issuer can have an automated speaker-verification system that maintains a data base of its card holders’ spoken phrases. The system enrolls each card holder by recording a set

of phrases spoken by the card holder to create models for identity verification. When a card holder requests access to his account, the automated system prompts the card holder to speak. The system then retrieves models of the stored spoken phrases. The newly acquired spoken phrases are compared with the retrieved models to determine if both were spoken by the same individual. As a result, a decision is made whether to grant the user’s credit account access request. In this example, both the enrollment and verification processes can be conducted by telephone. In many countries, telephones commonly are found in households and offices. Due to such wide availability of telephones, the data acquisition and transmission for speaker-recognition applications can be readily performed inexpensively.

“Speaker Recognition: A Tutorial” provides a road map for developing speaker-recognition algorithms. It describes each key component of a speaker-recognition algorithm: feature extraction and selection, pattern matching, and classification (the final decision). A simple example based on line spectrum pair frequency features using a divergence measure is described in detail.

—Weicheng Shen and Rajiv Khanna

Publisher Item Identifier S 0018-9219(97)06864-3.