Application of an Annular/Sphere Search Algorithm for Speaker Recognition

G. Miramontes de León, J. I. de la Rosa Vargas, and E. García Domínguez
Universidad Autónoma de Zacatecas
López Velarde 801, Zacatecas, Zac. 98064
(492)9239407 ext. 1518.

Abstract

In this work, an alternative search algorithm for vector quantization codebook is applied as a way to improve the performance of an automatic speaker recognition system. The search algorithm is based on geometrical properties of the vector space, defining annular and spherical regions instead of a full search method. The speaker recognition system is intended to identify a suspect, between a small group of persons, using low quality recordings, working as a text independent automatic speaker recognition system. Because the rate of recognition required in forensic applications is extremely important, the use of good discrimination algorithms can reduce the risk of bad decisions. The performance of the system under such a conditions is reported. Besides the few speaker samples used for training, a high recognition rate was obtained, so it was found an improvement of the recognition rate over the full search method.

1. Introduction

Automatic speaker recognition (ASR) is the process of recognize, automatically, a person based on the information included in a speech sample. ASR is one in a broad variety of biometric techniques. Other techniques are, for example, digital fingerprints, iris, DNA, etc. These techniques offer the possibility to control (accept or deny) the access to a variety of services like bank accounts, restricted areas in research labs, etc., but they also allows the identification of suspects in forensic applications, such as in threatening phonecalls from terrorist acts or kidnapping.

ASR can be classified in Automatic Speaker Verification (ASV), and Automatic Speaker Identification (ASI). ASV is the process of determining, among a registered number of speakers, the one who reclaims his identity, i.e., a person is saying to be someone and giving a speech sample of his voice, the process will find if there is a match, thus the system response will be *a*ccepted or *r*ejected. On the other hand, ASI is the process of finding between a group of persons the speaker under test. In addition, we can have a closed group ASI, if it is known a priori that the person belongs to the group, or open group ASI, if there is the possibility that the person under test doesn't belong to that group. This last classification is very important in forensic cases, since there is the possibility of an incorrect identification of a suspect. So it is very important to have improved decision methods. A false identification in this case will be extremely costly and has to be avoided always.

Because many decision methods are based on a minimal distance, for example, the Euclidian distance, and when the suspect (speech sample) doesn't belong to the group, the decision method depends on a threshold to avoid a false identification. This threshold is very difficult to determine and it is still an open problem. In this work, the main objective of the search algorithm is to restrict the search space so to reduce the risk of false identification, this is more important than reducing the time of the computations, even when, off course, this can happen.

Finally, there is *t*ext dependent ASR, when a fixed alphabet is used (the words to be used are known), and *t*ext independent ASR when the speech samples are randomly selected. Each case has its own field of applications, the former is valid for ASV, and the second is used in ASI.

The problem to be solved was the identification of a person having a low quality recording of his voice. In this case, the person under test denies to be the person who his voice was recorded. To complicate matters, there is no cooperation, so the study depends on a reduced number of speech samples and a text independent alphabet. Because of this, the study was conducted using only five different persons and only two sets of speech samples for each.

The following sections include: a brief description of the speaker recognition fundamentals, a description of the materials and methods employed during the tests, the results followed by some concluding remarks and future work.