

Voice disguise and automatic detection

Patrick Perrot ^{(*)(**)}, G. Aversano ^(*), G. Chollet ^(*)
perrot,aversano,chollet@tsi.enst.fr

^(*) CNRS-LTCL, GET-ENST, TSI Department

^(**) Institut de recherche criminelle de la gendarmerie nationale

Abstract

This study focuses on the question of voice disguise and the problem of its detection. The voice disguise is considered as a deliberated action of the speaker who wants to falsify or to conceal his identity. Lots of possibilities are offered to a speaker to change his voice and to false a human ear or an automatic system. He could transform his voice by electronic scrambling or more simply by exploiting the intra-speaker variability: modification of his own pitch, modification of the position of the articulators like lips or tongue which affect the formant frequencies. The proposed work is divided in three parts: the first one is a classification of the different possibilities available to change his voice, the second one presents a review of the different techniques used in the literature and the third one described the main clues proposed in the literature to distinguish a disguised voice from an original voice, before to propose some directions of research based on disordered and emotional speech.

Different means exist to change his voice : a classification of those techniques is proposed: electronic and non-electronic. The aim of this classification is to study separately each kind of disguise class and to determine some specific characteristics. A distinction is realized between electronic and non electronic conversion. In the case of electronic changes, the most sophisticated method consists in mimicking a specific voice. That is what is qualified as voice conversion. Different techniques of voice conversion are explored in the literature with more or less success. The main works on this field and their results are presented. The principle is to elaborate a conversion function between a source and a target voice and to apply it to the source voice. The aim is to obtain a transformed source voice that sounds like a target voice. The second class of change in the electronic field is what is qualified as voice transformation. It consists in modifying his voice by some specific methods. These methods could be separated in two categories: parametric methods based on an accurate signal model and non-parametric method based on temporal or frequency field.

The second main part of the classification is the non-electronic changes. In the field of voice conversion the different studies on the work of a professional impersonator are presented to understand the main features used to imitate a voice. At last in the register of non-electronic voice transformation, two categories is described: alteration of the voice by using a mechanic mean like a pen in the mouse for instance, and prosody alteration. This last category is very large because the impostors can modify lots of parameters of his voice. The position of the different articulators affect vowel sounds for instance, the use of a foreign accent changes some voice features, but also the modification of the pitch or the formants position and so on. Before to propose some specific parameters to study, a presentation of different works on the detection of voice disguise is presented. Most of the studies available in the literature concerns the most common voice disguise that is to say the prosody alteration. In order to organize our work some particular disguises has been chosen. Our aim is to be able to recognize automatically a disguise, to identify it and if it is possible to link the disguised voice to the original speaker. A description of the method that we plan in order to satisfy those different objectives is presented. The method is based for one part on the characterization of some specific features, like the pitch, the position of the formants... and for a second part on a clustering technique dedicated to elaborate specific models for each kind of disguise.