

Chapter 15

Pertinent Prosodic Features for Speaker Identification by Voice

Halim Sayoud
USTHB University, Algeria

Siham Ouamour
USTHB University, Algeria

ABSTRACT

Most existing systems of speaker recognition use “state of the art” acoustic features. However, many times one can only recognize a speaker by his or her prosodic features, especially by the accent. For this reason, the authors investigate some pertinent prosodic features that can be associated with other classic acoustic features, in order to improve the recognition accuracy. The authors have developed a new prosodic model using a modified LVQ (Learning Vector Quantization) algorithm, which is called MLVQ (Modified LVQ). This model is composed of three reduced prosodic features: the mean of the pitch, original duration, and low-frequency energy. Since these features are heterogeneous, a new optimized metric has been proposed that is called Optimized Distance for Heterogeneous Features (ODHEF). Tests of speaker identification are done on Arabic corpus because the NIST evaluations showed that speaker verification scores depend on the spoken language and that some of the worst scores were got for the Arabic language. Experimental results show good performances of the new prosodic approach.

INTRODUCTION

How can human beings recognize speakers only by their prosodic features (i.e., by their accents)?

In fact, a lot of information regarding the speaker identity is contained in his pitch, his duration and his energy: then in his prosody (Dehak, Dumouchel, & Kenny, 2007). In the task

of speaker recognition (Zhang & Tan, 2008), we are interested in identifying the speaker identity, whatever the used technique and used features, even if the recent research trends are to employ usual state-of-the-art features.

However, some works, using the prosodic features as the pitch contour, showed good performances in text-dependant speaker recognition. We can quote the first experiments of Atal in 1972 (Atal, 1972), who used a technique based on the

DOI: 10.4018/978-1-4666-0119-2.ch015

pitch contour with a learning time of 2 s. This technique provided a recognition score of 97% on small databases.

Since then, most of the research works have used cepstral features which seem to be more interesting in automatic speaker recognition, as a state of the art in this domain (Reynolds, 1995). In our opinion, that research trend is not very promising since it discourages any research work in another direction (other than the state of the art one). For that reason, we thought to investigate the effect of non-traditional techniques and proposed the use of three heterogeneous prosodic features for the task of speaker identification. In fact, many investigations, in the Intra and inter-Speaker variability on these features (Sayoud & Selmane, 1998) have been done, showing their relevance in speaker characterization.

For the choice of the classifier, we have proposed a modified LVQ1 algorithm (Lloyd, Brereton, Faria, & Duncan, 2007; Sayoud & Ouamour, 2004), which we called MLVQ1. We have also proposed a new metric which allows the association of the three heterogeneous features. We called it “ODHEF” or Optimized Distance for Heterogeneous Features.

DATABASE

We used an Arabic speech database which contains a lot of pertinent phonemes (a pertinent phoneme is a phoneme which characterize well the speakers) (Sayoud & Ouamour, 2004). These phonemes are given in table 1.

So, before the recording step, we asked the speakers to utter some sentences that are rich in some Arabic letters like: “Kaf”, “Rra”, “Ain” and “Dta” (see Table 1 for the Arabic correspondence).

This choice is made after many experiments on the speech variability (Sayoud & Ouamour, 2003): the speakers have been invited to utter some utterances of about 5.5 seconds in standard Arabic, so that each utterance is repeated 6 times.

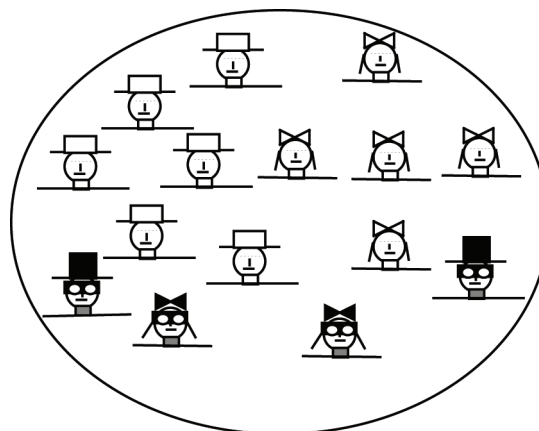
Table 1. Some pertinent Arabic phonemes used in our sentences

Phoneme designation	Latin similarity	Arabic Transcription
“Kaf”	K	ك
“Rra”	Spanish R	ر
“Ain”	Nothing	ع
“Dta”	Nothing	ط

We compose two sets of speakers; namely a closed set of 11 referenced speakers and an open one contained the 11 speakers considered as adherent and 4 other ones considered as impostors (Figure 1). In the closed set, there are 11 speakers voice: 6 males and 5 females (Figure 2). All the speakers are referenced by the system. However, in the open set we have some speakers who are not referenced (representing impostors).

Concerning the recording technique, the speech signal is recorded at a sampling frequency of 16 kHz with 16 bits of length and without noise.

Figure 1. The open set with 11 referenced speakers and 4 impostors: the masked speakers are impostors



13 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/pertinent-prosodic-features-speaker-identification/62975

Related Content

Design of a Hospital Interactive Wayfinding System: Designing for Malaysian Users

Ashok Sivaji, Hizbullah Kampo Radjo, Mohd-Faizal Aminand Mohd Azrin Hafizie Abu Hashim (2016). *Critical Socio-Technical Issues Surrounding Mobile Computing* (pp. 88-123).

www.irma-international.org/chapter/design-of-a-hospital-interactive-wayfinding-system/139560

Extended Mobile IPv6 Route Optimization for Mobile Networks in Local and Global Mobility Domain

Arun Prakash, Rajesh Verma, Rajeev Tripathiand Kshirasagar Naik (2010). *International Journal of Mobile Computing and Multimedia Communications* (pp. 1-17).

www.irma-international.org/article/extended-mobile-ipv6-route-optimization/43890

Next-Generation Mobile Technologies

C. Tan, C. Chinand M. Sim (2007). *Encyclopedia of Mobile Computing and Commerce* (pp. 700-705).

www.irma-international.org/chapter/next-generation-mobile-technologies/17159

MIMO-Based Secret Key Generation Strategies: Rate Analysis

Kan Chenand Bala Natarajan (2014). *International Journal of Mobile Computing and Multimedia Communications* (pp. 22-55).

www.irma-international.org/article/mimo-based-secret-key-generation-strategies/130480

How Do People Use Their Mobile Phones?: A Field Study of Small Device Users

Tianyi Chen, Simon Harperand Yeliz Yesilada (2011). *International Journal of Mobile Human Computer Interaction* (pp. 37-54).

www.irma-international.org/article/people-use-their-mobile-phones/51656