Automatic Spoken Language Identification by Digital Signal Processing Methods. Tatar and Russian Languages

Rustam Latypov[(✉)], Ruslan Nigmatullin, and Evgeni Stolov

Kazan Federal University, Kazan 420008, Russia
`roustam.latypov@kpfu.ru`

Abstract. The paper studies the problem of language identification for audio files. For solving the problem, we use methods of digital signal processing only (without analysis of phonemes distinctive for language). A special attention is drawn to the form of signal in an area close to the position of a stop consonant. The evaluation is performed on a set of two languages; this includes speech records taken from TV programs. It is provided that solely one of the two languages (either Tatar or Russian) is used in each of files. Experimental evidence demonstrates the feasibility of the proposed techniques.

Keywords: Language identification · Tatar and Russian languages · Form of signal

1 Introduction

Automatic spoken language identification (LID) is a process by which the language spoken in a digitized speech sample is recognized by a computer [1]. Applications of LID systems include front-end ones for speech recognition, automated dialogue systems, call routing, call centers, household devices, wiretapping, and information distillation. Research in automatic LID systems possesses a history going back to the early 1970s: Doddington and Leonard [2] have studied frequency of occurrences of certain reference sound units in different languages. Discriminations between languages can be determined by employing paralinguistic information in a speech signal including distinctions in the phonology, morphology, syntax, and prosody [3]. In recent years, great improvements were achieved in different LID fields of research. The most popular modeling techniques used in acoustic systems are in application of some spectral features such as mel frequency cepstral coefficients (MFCC) [5]. In contrast, phoneme n-gram statistics are modeled in order to recognize languages in phonotactic approaches [6, 7]. Note also a widespread use of deep neural networks [8] as well as linear prediction technique [9]. However, such LID systems require greater computing resources to run. The latter narrows the scope of their applications (for example, in online systems or systems without Internet connection). In our work, we search the new speech parameters in LID to reduce computational complexity of LID algorithms. The speech file is treated as a digital signal and recognizing the language of a file is based on digital signal processing methods. The Tatar and Russian are used in the capacity of