Sound Classification based on Feature Extraction

Martin MOKRÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`xmokry@fiit.stuba.sk`

Voice control of a computer is relatively new concept. It is no surprise, that it is becoming very popular, considering the fact, that it is the most comfortable and effective way of communication between people. It has become a subject of study for many researchers in past years. However real life applications use mostly words, which are quite long voice commands and executing more commands in a row can feel tedious.

Our work focus on application control by short articulate (e.g. syllable) and inarticulate (e.g. clap, whistle) sounds. This makes application control more effective, although not very intuitive at start.

Real life application usage creates problems, we need to be prepared to deal with. The most significant ones are [1]: environment noises (e.g. people talking in the background) and different speakers (e.g. dependence on gender or age). Because of these problems, the proposed approach needs a data representation, which has low sensitivity to noise and allows slight variations between the tested sequence and the template. That is why, we decided to find similarities between sounds through sound features, which have been proven to perform well in a noisy environment and also have small memory requirements [2].

Features extracted from the recorded sound during test phase need to be classified in real-time in order to ensure quick response. For this purpose we chose to test two classifiers: Naive Bayes, k-Nearest Neighbors.

We believe that high accuracy of real time classification of articulate and inarticulate sounds can be achieved by following these steps:

1. Segmentation - division into the windows of the same length.

2. Feature extraction - calculation of characteristic values of sound for each window.

3. Feature normalization - transformation of real values into intervals and normalization using z-score.

---

* Supervisor: Jakub Ševcech, Institute of Informatics, Information Systems and Software Engineering

4. Classification - class determination for every window, followed by labeling of the whole sound by class name.
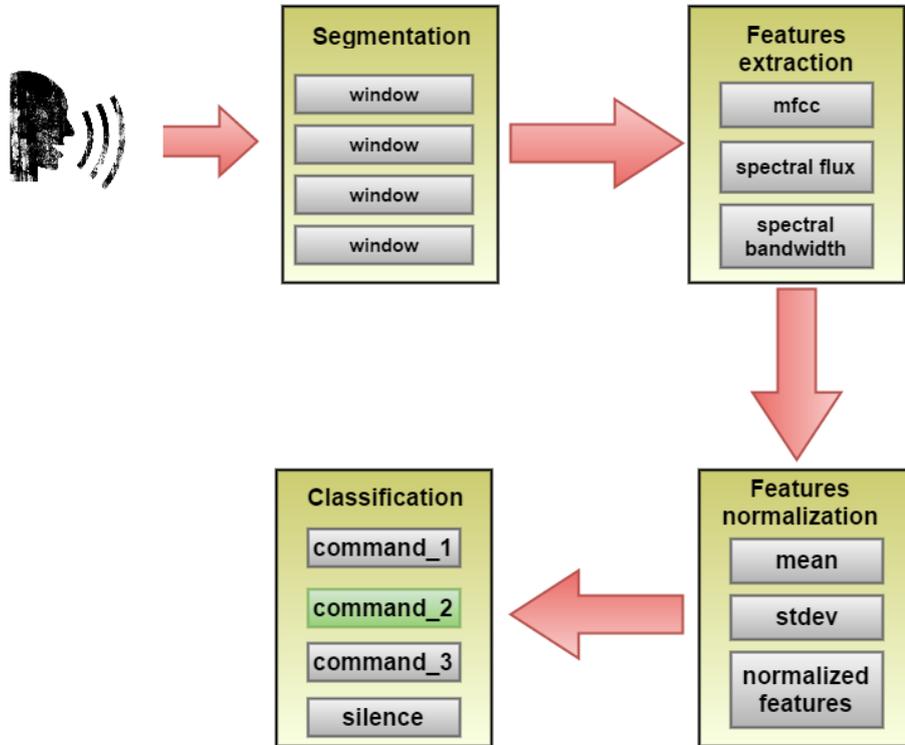

Figure 1. Illustration of steps of proposed method.

Extended version was published in Proc. of the 12th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2016), STU Bratislava.

References

[1] Anusuya, M., Katti, S.: *Speech recognition by machine: A review,* International Journal of Computer Science and Information Security, (2009)

[2] Nanopoulus, A., Alcock R., Manolopolus, Y.: *Feature-based classification of time-series data,* Information processing and Management, (2001)