BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

37

Excel@FIT 2018

Implementation of a simple speech recognition engine for Android

Eduard Čuba
Brno University of Technology, Faculty of information technology

Motivation

Lack of customizable speech recognition frameworks, optimized for mobile devices, which might be easily integrated in other applications.

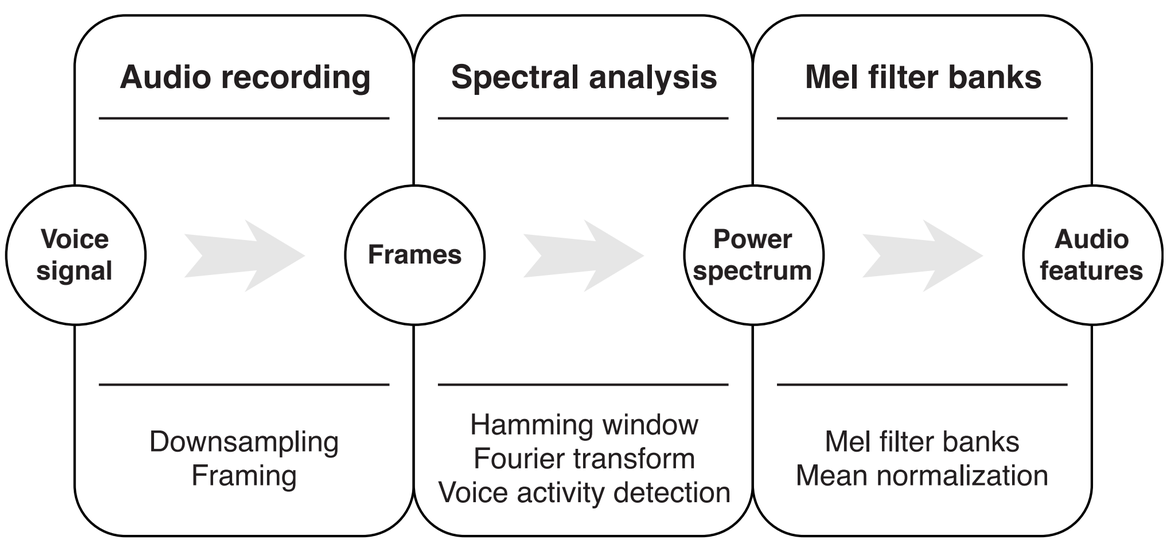
Problem

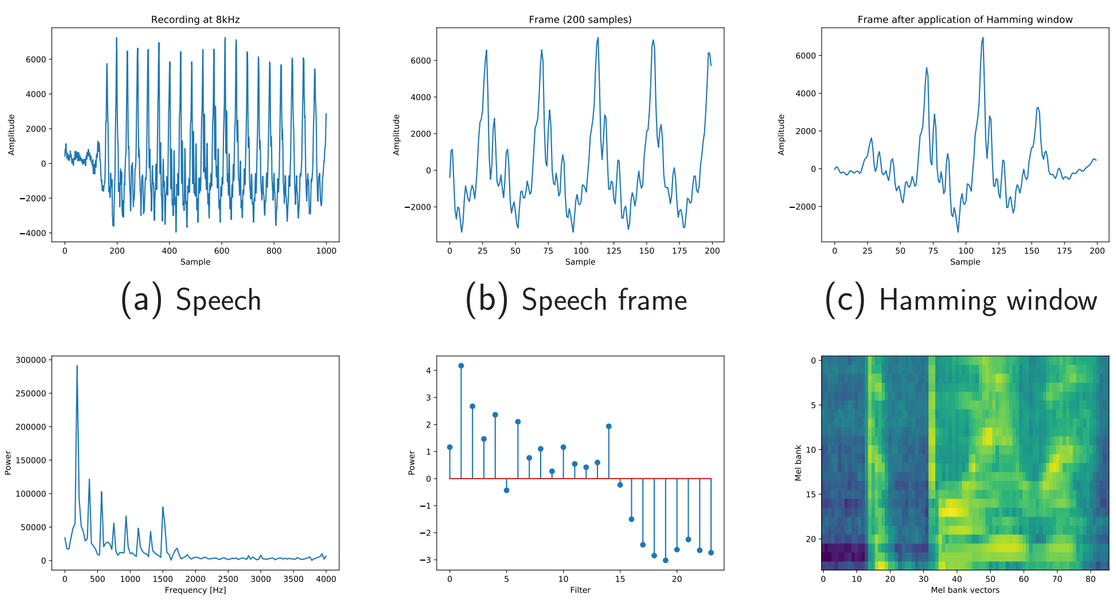
Insufficient resources of mobile devices for running conventional speech recognizers in real-time and high memory requirements of static decoders.

Goal

To create an easy-to-use library, optimized for mobile devices, which would provide a speech recognition based on custom models trained for the desired application.

1. Acoustic feature extraction





(a) Speech (b) Speech frame (c) Hamming window

(d) Power spectrum (e) Mel filter banks (f) Acoustic features

2. Posterior phoneme probability estimation

Single state posterior phoneme probabilities are estimated by an artificial neural network with a bottle-neck layer. We built a system for parallelized computation of phoneme posteriors using appropriate vectorization techniques (e.g. NEON) for mobile architectures, with an emphasis on real-time performance.

3. Acoustic modeling

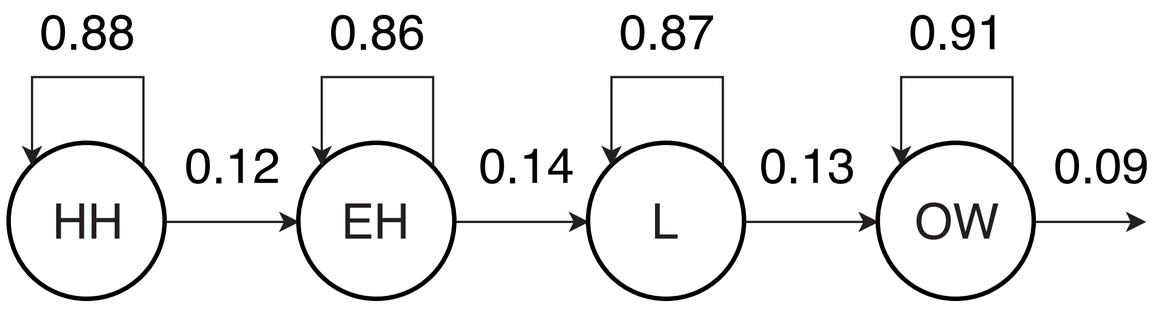


Figure 2. Acoustic representation of the word "Hello". The phoneme tag inside the transducer shows the phoneme, which prior probability should be used. The value above the transition arrow defines a probability of leaving the transducer.

4. Dynamic decoder implementation

We tried to overcome the high memory requirements of static decoders by using a dynamic one based on small uni-gram recognition network. N-gram probabilities are applied at run-time by performing a lookup in an appropriately designed data structure.

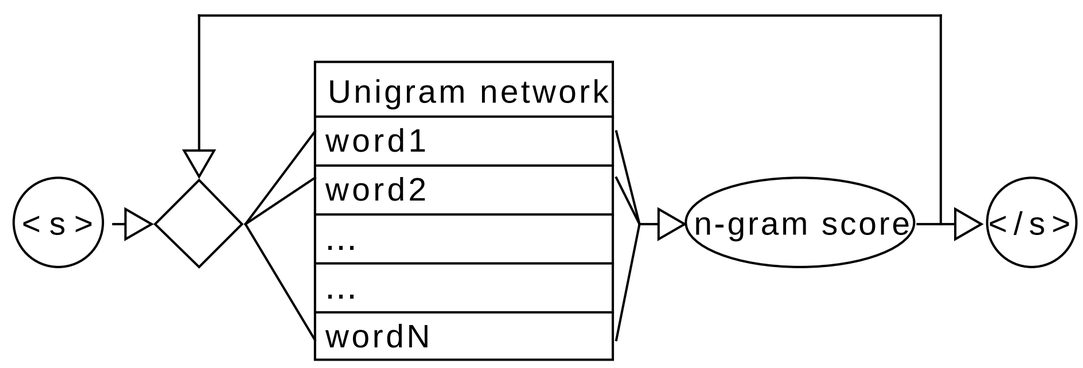


Figure 3. Sample representation of a dynamic decoder with uni-gram network and dynamic *n*-gram probability application.