Porticus Speaker Identification System for

NIST 2010 Speaker Recognition Evaluation

PORTICUS TECHNOLOGY

Algirdas Bastys : algirdas.bastys@mif.vu.lt
Bernardas Salna : bernardas@centras.lt

June 2010

System description

Preprocessing, features

For initial preprocessing the all NIST SRE 2010 data were filtered with eight sub-band filters with central frequencyranging from 2000 Hz to 4000 Hz. The complex valued filters were defined by the following formula:

,

where and , . The filter has only one parameter having transparent interpretation and allows an efficient implementation requiring about 10 arithmetic operations per sample.

The filtered datawere divided into 0.02 sec. duration and 50% overlapping frames. Before estimation of features of a frame we transformed complex valueddata to real valued data by centering the filtered data along maximum of their magnitude. More exactly, ifandare start and end time moments of the n-th frame, than the frame features were extracted for the real-valued function

, ,

where , and .

For each fixed sub-band index *k* and frame number *n* twenty (ignoring the zeroth, as usual) Linear Predictive Cepstral Coefficients (LPCC) of windowedfunction were calculated. Eight symmetric covariance matrices , , of the sequences of the LPCC frames features were calculated. Thus the total number features of a speech utterance was 8x20x(20+1)/2 = 1680.

Distance measure

Comparison of two speech utterances is done by means of arithmetic-harmonic sphericity measure. Letand , , are covariance matrices of a train and test model respectively. Than distance measure between the two models is defined by the following formula:

,

where and are respectively arithmetic and harmonic means of eigenvalues of matrix. The distance measure is non-negative and equals to zero iff the all eight covariance matrices of the two models are proportional.
Z-normalization of distance measures were performed using NIST SRE 2008 short2-short3 data. Final results were presented in LLR form.

Results

Time consumptions for creation of train and tests models are symmetric; using a standard Laptop the all NIST sre 2010 data were processed in about 10 hours or 100 sec. of raw data handled in 1 sec. For classification of the all presented core pairs we used about 20 min. of the laptop processor time.

Nr	Train/Test conditions	EER (%)
1	Interview in test and train, same microphone	12
2	Interview in test and train, different microphone	26
3	Interview versus phone call, channel tel	28
4	Interview versus phone call, channel microphone	19
5	Phone call in train and test, different number	30
6	Phone call in train and Hve phone call in test, channel tel	37
7	Phone call in train and Hve phone call in test, channel mic	28
8	Phone call in train and Lve phone call in test, channel tel	21

9	Phone call in train and Lve phone call in test, channel mic	16

Table 1. Equal Error Rate (EER) for different NIST sre 2010 core-core conditions.