LEHRSTUHL FÜR
MUSTER-
ERKENNUNG

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Lehrstuhl für Mustererkennung (Department Informatik)

Šárka Nesvedová
Faculty of Information Technology
Brno University of Technology
Božetěchova 2
612 66 Brno
Czech Republic

Computer Science Department 5
Pattern Recognition Lab
Informatik 5
Lehrstuhl für Mustererkennung
Prof. Dr.-Ing. Elmar Nöth
Martensstraße 3, 91058 Erlangen
Room 09.136
Telephone +49 9131 85-27888
Fax +49 9131 303811
noeth@cs.fau.de

Your reference 18/1493/20102
Your message from 17.2.2012
Our reference Kockm01

Erlangen, March 24, 2012

Expert opinion on the doctoral thesis

“**Subspace Modeling of Prosodic Features for Speaker Verification**”

of Dipl. Ing. Marcel Kockmann.

The thesis deals with the subject of speaker identification/verification. This subject is of high interest in many different areas with applications ranging from access control in telephone banking to intelligence applications. In the latter application which is in the focus of this thesis few assumptions can be made w.r.t. the content of the speech signal, the transmission channel and even the spoken language. As a consequence, most approaches in the field rely on very low level acoustic information. Mr. Kockmann’s research concentrates on the use of prosodic information, robustness towards channel mismatch and combination of the prosodic information with acoustic information. Due to the high interest in the research field, there are regular competitive evaluations of speaker verification systems so that Mr. Kockmann is in the comfortable situation that he can work on well established data sets and compare and combine his work with other state-of-the-art systems.

The thesis is structured into 6 chapters.

Chapter 1 gives a short overview over the development of speaker verification systems in the context of the NIST evaluations in the last 20 years. The principle approach of the systems is explained.

Chapter 2 introduces the data and evaluation measures that are used in this thesis and in several NIST evaluations.
Chapter 3 describes the parameterization of the prosodic features. The speech signal is segmented into syllable-like units and the fundamental frequency and energy within these units is described with DCT-coefficients. Alternatively, the prosodic information is described with the so-called SNERF features which were developed at SRI. Some results are presented that show the influence of different segmentation methods and exactness of the contour approximations.
Chapter 4 describes the modeling approach of Mr. Kockmann's system, starting from a standard GMM-UBM system, adding the JFA and subspace models. Experimental results confirm the improvements of the additional processing steps.
Chapter 5 is the main experimental chapter, where the individual prosodic systems, the baseline acoustic system (based on cepstral features) and the fusion systems are evaluated.
Chapter 6 summarizes the thesis and points to future work.

The work of Marcel Kockmann clearly fulfills the requirements of a doctoral thesis:

- The topic is appropriate to the area of dissertation and is up-to-date w.r.t. the present level of knowledge.
- The work is original and is an important contribution to the research field of speaker verification. The main contributions are
  - Fixed-size parametric representation of the fundamental frequency and energy contour based on a robust segmentation into syllable-like units.
  - Making these prosodic features robust to channel variation by using state-of-the-art techniques such as Joint Factor Analysis (JFA) and subspace models.
  - Applying the JFA to the SNERF features from SRI.

  - Fusing his own approach with the SNERF approach on a feature level.
  - Fusing the prosodic and acoustic information on the feature level.
  - Showing that prosodic information produces worse results than acoustic information (what is to be expected), but that the prosodic information is complementary and thus that the fusion of both information sources clearly surpasses the results of state-of-the-art acoustic systems.
- His results are directly comparable with the results of the best groups worldwide, because he evaluates on publicly available standard databases.
- His publication list is impressive for a young researcher including contributions to NIST speaker evaluation workshops, ICASSP, Interspeech, and Speech Communication. At Interspeech 2009 he successfully applied the technology to speaker independent emotion classification and won the 5-class-emotion-challenge.
- The overview over the state-of-the-art and the cited literature demonstrate his thorough knowledge of the speaker verification field.

To summarize, the doctoral thesis clearly meets the requirements of the proceedings leading to the PhD title conferment.

Elmar Nöth

Prof. Dr.-Ing. Elmar Nöth