



I4U submission to NIST SRE 2012: A large-scale collaborative effort for noise-robust speaker verification

R. Saeidi¹, K. A. Lee², T. Kinnunen³, T. Hasan⁴, B. Fauve⁵, P. -M. Bousquet⁶, E. Khoury⁷,
P. L. Sordo Martinez⁸, J. M. K. Kua⁹, C. H. You², H. Sun², A. Larcher², P. Rajan³, V. Hautamäki³,
C. Hanilci³, B. Braithwaite³, R. Gonzales-Hautamäki³, S. O. Sadjadi⁴, G. Liu⁴, H. Boril⁴, N. Shokouhi⁴,
D. Matrouf⁶, L. El Shafey⁷, P. Mowlae¹, J. Epps⁹, T. Thiruvaran⁹, D. A. van Leeuwen¹, B. Ma², H. Li²,
J. H. L. Hansen⁴, J. -F. Bonastre⁶, S. Marcel⁷, J. Mason⁸, E. Ambikairajah⁹

¹Radboud University Nijmegen, The Netherlands, ²Institute for Infocomm Research, Singapore,

³University of Eastern Finland, Finland, ⁴CRSS, University of Texas at Dallas, USA,

⁵ValidSoft Ltd, London, UK, ⁶LIA, University of Avignon, France

⁷Idiap Research Institute, Switzerland, ⁸Swansea University, UK,

⁹University of New South Wales, Australia

Abstract

I4U is a joint entry of nine research Institutes and Universities across 4 continents to NIST SRE 2012. It started with a brief discussion during the Odyssey 2012 workshop in Singapore. An online discussion group was soon set up, providing a discussion platform for different issues surrounding NIST SRE'12. Noisy test segments, uneven multi-session training, variable enrollment duration, and the issue of open-set identification were actively discussed leading to various solutions integrated to the I4U submission. The joint submission and several of its 17 sub-systems were among top-performing systems. We summarize the lessons learnt from this large-scale effort.

Index Terms: Speaker Verification, NIST SRE 2012, I4U, i-vector

1. Introduction

The I4U submission to National Institute of Standards and Technology (NIST) speaker recognition evaluation 2012 (SRE'12) [1] is a result of active exchange of information between the coalition participants across nine institutions. The name of the institutes and corresponding system identifiers are provided in Table 1. The submitted results are based on the fusion of multiple classifiers. The optimization of the component classifiers and the fusion device were done with development sets jointly designed within the I4U coalition with multiple design iterations, refinement of noise adding protocol and various other details. Different from previous SREs, the task of SRE'12 involves:

Handling noisy test segments: This required speech enhancement algorithms and employing mixed training or parallel model combination techniques.

Imbalanced multi-session training: There are tens of segments available for training some speaker models while only a single segment for some other speakers.

Open-set identification: SRE'12 evaluation protocol allows the use of knowledge of all target speakers in each detection trials which resulted in utilizing compound log-likelihood ratio.

This work was partly supported by Academy of Finland (proj. no 253120 and 132129), Swiss National Science Foundation under the LOBI project, contract no. SNSF-235 and European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no. 238803.

Table 1: I4U Coalition and assigned system indexes

Site	System index
ValidSoft Ltd (VLD)	Sys1
Swansea University (UWS)	Sys2
University of Avignon (LIA)	Sys3
Radboud University Nijmegen (RUN)	Sys4
University of Texas at Dallas (CRSS)	Sys5-10
University of Eastern Finland (UEF)	Sys11
Institute for Infocomm Research (IIR)	Sys12-16
Idiap Research Institute (IDIAP)	Sys17

This paper is organized as follows: In Section 2, we present the strategies taken to make a development set coping with SRE'12 new conditions. Details of the submitted systems and the component classifiers, together with the strategies to deal with the new challenges listed above are described in Section 3. One of the motivations underlying the I4U coalition is to experiment with the fusion of large numbers of sub-systems. Results for the individual and the fused system are presented in Section 4.

2. Development sets

The development sets were generated to help I4U team members in developing their speaker recognition systems considering the *special* conditions in SRE'12 including multiple segments training for a speaker¹. All the members of I4U coalition helped in refining the lists with respect to detecting empty or otherwise problematic segments with conflicts in gender and speaker PIN (there are issues with pre-SRE'12 lists like multiple-genders or wrong genders for some speakers). The latest lists from NIST were utilized and speech segments for all 1918 target speakers were fetched from SRE'06, SRE'08 and SRE'10 corpora and corresponding meta-data were extracted. To be able to assess both the recognition systems' generalization and calibration performance, separate development (DEV) and evaluation (EVAL) sets were created. The number of segments, speakers and trials for each set are given in Table 2. In designing these sets, the following criteria were considered:

- Test segments are disjoint for DEV-test and EVAL-test.

¹The lists are available via http://cls.ru.nl/~saeidi/file_library/I4U.tgz