

ALBAYZIN-08 System Evaluation Plan

Language Verification

(ALBAYZIN-08 VL)

Aim

The main goal of this evaluation is to promote the exchange of ideas, to stimulate creativity, to favor the collaboration between research teams that focus their research on Language Identification and Verification. With this purpose in mind, we propose a contest among systems working in a language verification task, similar to the one organized by NIST but with less difficulty: there are only 4 target languages (Spanish, Catalan, Basque and Galician) and the signals have more bandwidth and SNR than in NIST evaluation.

General conditions

The participating teams will have three months to develop their systems, using the specific training and development materials that will be provided. They will be able to use any other material and subsystems (speech recognizers, acoustic-phonetic decoders, etc.) that will have to be mentioned in the system description, including the corresponding references. Next, an evaluation corpus will be provided and the research teams will have three weeks to process it and send back the results. Each test file will be verified against the 4 target languages and it is possible that more than one language (or none of them) are detected simultaneously. A perfect system would detect only the actual language of each test file or no language if the file contains an unknown language.

The training, development and test materials are taken from TV programs (news, documentaries, debates, interviews, reports, magazines, etc.) from three independent and non overlapping sets. Although part of the signals have been recorded with relatively difficult signal and channel conditions (interviews in the street, with music, noise, etc.) most of them are clean study recordings. All the signals have been recorded using a digital Roland Edirol R-09 recorder and have been sampled at 16 KHz, 16 bit per sample and saved in mono Windows PCM wav files.

Verification results will be evaluated with the script from *2007 Language Recognition NIST Evaluation*, previously adapted to the proposed task. This means that results must be provided in NIST format: a text file with a row by each verification test and 6 text fields in each row, referring to the competition, the target language, the operation mode (see below

Competition conditions), the test file name, the decision made with respect to the detected language and the score assigned by the system. The bigger this score is the bigger is the probability of detecting the target language. Results will be expressed using DET curves that will show the minimum cost and working points. When the score is a *Log-Likelihood Ratio* (LLR), the average value of C_{LLR} function will also be provided.

Evaluation corpus

The evaluation corpus, of no more than 8 hours, will have no more than 2000 files in the 4 target and other unknown languages randomly mixed and with random names. The evaluation corpus will contain signals of three different lengths (3, 10 and 30 seconds). In any case, each signal will mainly contain speech and only some segments with silence or noise.

Training and development material

The training corpus will contain approximately 8 hours for each language (more than 30 hours in total) in files with different length. Similarly to the test materials, they will mainly contain speech and only some segments with silence or noise. Together with this training corpus, a development corpus similar to the evaluation corpus will also be supplied. The evaluation script will also be provided so that each research team can evaluate the systems they will develop.

Competition conditions

The systems can operate in open and closed mode. This determines two types of competition. In closed mode, the test signals must correspond to one of the target languages. In open mode, the test signals can correspond also to unknown languages. Concerning the evaluation in both cases the same test signals will be provided, but in closed mode the signals corresponding to unknown languages will not be considered.

On the other hand, taking into account the materials used to train the systems, another two competition types will be distinguished: free development (systems are free to use any data and subsystem) and restricted development (only the material provided in this competition can be used).

Each group can send 4 primary systems, corresponding to the following competition conditions: closed mode / free development, closed mode / restricted development, open mode / free development and open mode / restricted development. Besides, they can send as many alternative systems as desired. In the competition only the results from the primary systems will be considered. In all cases the results will be disaggregated by signal length.

Provisional schedule

- May 5th , 2008
 - Publication of the detailed evaluation plan in the JTH2008 web
 - Beginning of the registration term
 - Registration form available in the JTH2008 web.
- June 30th ,2008
 - Evaluation script and development and training materials released
 - Building of a collaborative WEB to submit doubts, make comments, etc. with a public part and another part only for the participants
- July 15th , 2008
 - Deadline for registration
- September 29th , 2008
 - Test corpus released
- October 19th , 2008 (till 24:00 GMT+2)
 - Deadline for submitting the verification results, following the established method and format in the detailed evaluation plan
- October 31st , 2008
 - Notification of the evaluation results to each of the participants
 - Test corpus language labels released

Registration

The research teams willing to participate in the evaluation campaign will have to fill in the form available in JTH2008 web since May 5th 2008. The registration deadline is the July 15th 2008. Before registering, it is recommended to read carefully the detailed evaluation plan that will be available in the JTH2008 web since May 5th. This plan might suffer modifications, due to restrictions, needs not foreseen initially or errors, which in any case will be communicated to all the participating groups. The interested ones can go directly to the organizing committee to consult aspects that are not clear in the evaluation, by e-mail to luisjavier.rodriquez@ehu.es and by phone (946012716).

Commitments of participating teams

1. To use the data provided exclusively with research aims. Participants will be able to use the database to develop or evaluate their own systems, but they will not be allowed to distribute nor sell them. They will have to reference the database as follows:

KALAKA. Speech database created for the 2008 Language Recognition Evaluation on Spanish Languages, organized by the Spanish Network on Speech Technology. Produced by the Software Technologies Working Group (GTTS, <http://gtts.ehu.es>), University of the Basque Country.

2. Each participant will have to submit at least the results of one of the primary systems. The sending procedure will be specified in the detailed evaluation plan. In any case, it will have to contain the team identification and a results file by each system, indicating whether it corresponds to a free or restricted system, whether the system is operating in open or closed mode and whether the scores are or not *log-likelihood ratios*.
3. Each participant will be expected to submit in addition to the synthesized waves, a paper (preferably in format PDF) describing the main characteristics of their system (or systems). These contributions will be distributed to all the participants together with the final report of the organizing committee of the ALBAYZIN-08 TTS evaluation in the V Workshop on Speech Technology. In order to homogenize the aspect of the papers, please use the Latex or MS-WORD templates available in the JTH2008 web.
4. To be present at the V Workshop on Speech Technology (JTH2008), to be held on November 12-14th 2008, to explain the technical details of their language verification system.