

EVALITA SPEECH TASKS

Franco Cutugno¹ and Mauro Falcone²

¹Università Federico II di Napoli

²Fondazione Ugo Bordoni

The history of organized collaboration in speech system evaluation has born more than 20 years ago. In Europe it started with COST projects, and then it continued with a plethora of ESPRIT and other European Projects. In the United States after the pioneering DARPA activities, the successful NIST model was unanimously recognized as the winning approach as it combines the definition of standards and measures, the collection of databases, the organization of open evaluation campaigns, as well as (last but not least) the definition of new and interesting tasks and objectives which are yearly updated. The NIST initiatives were able to define a kind of “ecosystem” into which researchers and companies may enter in and expertise a sane competition aiming at improving the know-how and the potentialities of all them.

As a matter of fact, despite of the growth of the successes of the above cited projects, an institution having the same aims of NIST, has never seen the light in Europe. In this frame a further dichotomy is due to the apparent division existing, especially in Italy, between “textual” computational linguistics and speech sciences. It is not a real competition, as one could expect, for funds- or visibility- hunting, but, instead, it is a sort of reciprocal indifference.

A further and non minor problem is then represented by the fuzzy status of speech technology Italian community which is almost present and active both in ISCA (with about the same number of members than the prestigious Sweden) and in European Projects where Italy is widely represented by companies and scientists, but completely fragmented and often in contrast with phoneticians and linguists when the community is considered under an internal, country-related view.

Two years ago a ‘pioneer’ (respect to the Italian situation) group of researchers attempted for the first time to organize an evaluation campaign leading to the first edition of EVALITA. At that time organizers decided of not involving in the initiative speech scientists.

Speech task evaluation has been then introduced in the second EVALITA edition, in an already existing “text-evaluation” framework thanks to the obstinacy of the authors of this introduction, with the sponsorship of the Italian Association of Speech Science (AISV). Since the first days of work on this project, we were aware that this was a risky project; nevertheless considering our past experiences in speech evaluation, database collection and our sincere passion in speech technology tout-court, and in spite of a rationale evaluation of odds and risks of success of the project, we decide to go ahead. When we officially started with the EVALITA speech task, the organization of the “text” events was already completed and we only had a broad idea of which task and persons we may involve in this adventure. In addition we also had the problem to find all the speech resources in Italian language, and this is not a trivial problem given the status of speech research in Italy. On one side, we were

lucky and we find a very good pool of organizers for the different speech tasks and tracks. If nowadays we can say that the EVALITA'09 speech tasks led to a successful experience, we had to thank the task organizers: (Guido Aversano, Niko Brummer, Gianpaolo Coro, Morena Danieli, Roberto Gretter, Marco Matassoni, Roberto Pieraccini, Giuseppe Riccardi, Luciano Romito) that really did a great job.

We defined, for this first experience, three tasks in speech input technology: digits recognition, dialogue systems, and speaker verification. Digits recognition includes two tracks (clean and noise condition), speaker identity verification is also divided into two tracks (application and forensic). These evaluations were carried on according the well known evaluation methodologies described in the scientific literature. Dialogue systems have been evaluated on the basis of an expert trial and evaluation procedure. In total we had 17 different participants in speech tasks. We learn a lot from this experience and we believe that it could be repeated in a wide multilingual pan-European scenario, including, of course, Italian. As a final commitment, new tasks, related to new research field and emerging applications, must be thought for the future evaluation campaigns.