

# SpLC: Speaker and Language Characterization (SpLC): A Special Interest Group (SIG) of ISCA

*Jean-François Bonastre (1), Ivan Magrin-Chagnolleau (1), Stephan Euler (2)  
François Pellegrino (3), Régine André-Obrecht (4)  
John Mason (5), Frédéric Bimbot (6)*

- (1) LIA University of Avignon, Avignon (France)
- (2) Robert-Bosch GmbH, Stuttgart (Germany)
- (3) DDL, University of Lyon II, Lyon (France)
- (4) IRIT, University Paul Sabatier, Toulouse (France)
- (5) University of Swansea, Swansea (UK)
- (6) IRISA/INRIA, Rennes (France)

## Abstract

Last year, SpLC - an ISCA Special Interest Group (SIG) centered around Speaker and Language Characterization was born. The aims of this paper are to present the SpLC SIG, its objectives, and the work done during the first year.

## 1. Introduction

The SpLC ISCA-SIG was launched during the RLA2C workshop, in 1998, with the objectives of providing an entry point and a network for all the persons involved in the field of Speaker and Language Characterization, in order to facilitate the exchange of information, to stimulate discussions between the different communities and countries, to compare viewpoints and approaches, etc.

People interested in speaker and language characterization have various origins: computer science, computer engineering, electrical engineering, speech science, phonetics, linguistics, etc. This characteristic highlights the need of collaboration and exchange of information.

The aims of SpLC are:

- to facilitate exchanges between members;
- to provide information on the area;
- to promote SpLC activities;
- to provide members of ISCA with a special interest in SpLC with a means of exchanging news of recent research developments and other matters of interest in SpLC;
- to sponsor meetings and workshops in SpLC that appear to be timely and worthwhile, operating within the framework of ISCA's by-laws for SIGs;
- to provide and make available resources relevant to SpLC, including text and speech corpora, analysis

tools, analysis and generation software, research papers and generated data.

## 2. SPLC Internal Organization

SpLC organizes its activities around four main areas:

- Academic research on Speaker Characterization and Recognition;
- Forensic aspects of Speaker Characterization and Recognition;
- Language Characterization and Recognition;
- Commercial Applications for Speaker and Language Recognition.

In addition, several transversal activities will be proposed. For the moment, three are already defined:

- *Evaluation group*: this group works on the evaluation paradigm, for all the various axes of the SpLC domains. Particularly, the group will highlight evaluation tasks for commercial applications, for speaker tracking and indexing, and for language identification.
- *Meetings and Workshops group*: this group concerns the organization of SpLC-events, like workshops, meetings, summer schools, exhibitions, etc.
- *Speaker Indexing group*: this group proposes to work on algorithms to do speaker indexing, which is a recent task in the speaker recognition community.

## 3. The main areas

As written before, SPLC is organized around four main areas. Each main activity has (or will have) its own group leader(s) and its electronic communication tools (at the beginning, one electronic mailing list). Each activity organizes the work, the meetings, etc.

### 3.1. Academic research on Speaker Characterization and Recognition

Automatic speaker recognition is just one kind of biometric security mode. It is deemed not to be particularly accurate in relation to some other modes, but is very convenient. Thus increasing the accuracy of the speech mode has been the goal of many researchers for many years. The recent NIST open trials has focused the work of several leading sites, with controlled, inherently competitive evaluations. This has unquestionably helped towards achieving the goal of greater accuracy, largely it would seem through the introduction of new normalization techniques.

Speaker recognition accuracy in practice is constrained by natural variations in an individual's speech from one occasion to another and the recording conditions. Both cases demand capturing sufficient representative training data, since strategies for generalization are currently inadequate. Strategies to overcome or circumvent these limitations are topics for today's research.

One potentially productive avenue for improved accuracy is multi-modal biometrics. The suitability of different modes is likely to be very application dependent, though obvious candidates to complement speech (audio) are visual speech based on lips and the face generally [1],[2],[3].

Possible roles for SpLC here are clear:

- to foster multi-modal work in biometric security where speech is an integral part.
- to evolve new groupings that identify and publicise on-going research in multi-modal biometrics.

A good example of the latter is AVBPA (Audio- and Video-based Biometric Person Authentication). In the Association's own words:



“AVBPA has an ambition to bring together the leading biometric signal analysis researchers of image and voice modalities, in an attempt to contribute with robust solutions to efficient and secure communication. The conference, that has been organized biannually, is an official event of the International Association of Pattern Recognition”

The third biannual conference is to be held in Halmstad, Sweden in June 2001.

As mentioned above, speech is deemed not to be one of the more accurate of modalities. However it does have one distinct advantage over most other biometric modes, namely the ease with which an identity claim can be made; for this can be a fully integrated part of the authentication process. Thus

speech is an ideal partner to other modalities. Consequently, SpLC activities can look towards the example of AVBPA and extend current activities in multi-modal work.

In so doing it might well be worth considering the different modes under the biometric classification of Roethenbaugh and Mansfield [4]. In their revised glossary of terminology emphasis is given to *physiological* and *behavioural* biometrics. Essentially, physiological biometric is what you *are*, while a behavioural biometric is what you *do*. Some examples are given in the table below.

	Physiological What you <i>are</i>	Behavioural What you <i>do</i>
Eye-related scan	are	
Fingerprints	are	
Speech(audio)		do
Handwriting		do
Face 	are	do
Lips 	are	do

Thus speech provides an interesting dual biometric

### 3.2. Forensic aspects of Speaker Characterization and Recognition

Forensic applications of speaker characterization and recognition occupy a special position in the field of speech processing.

Firstly, because forensic applications do not have only a financial impact, but also a social one, given their potential consequences on the life of a person. Secondly, because the forensic field involves a wide range of people, i.e. researchers, industrialists, lawyers, judges, investigators, policemen, and sometimes even jury members, that is, anybody. This requires efforts to improve communication between this variety of people.

A lot of issues related to forensic speaker recognition remain open. Is it scientifically possible to identify someone using his/her voice? Is it possible to provide some robust information in order to help investigations? How can the reliability of a forensic technique be assessed?

The SpLC is composed of several groups, among which the “forensic group”. An important activity of the forensic group could be to make a description of the judicial systems in various countries all over the world, and to compare the place of voice analysis and speaker recognition in the different systems.

### 3.3. Language Characterization and Recognition

Automatic Language Identification appeared in the United States 25 years ago; the 90's saw the expansion of this theme, when the first multilingual corpora were made available. Beyond the interesting linguistic point of view, significant stakes lie in the Multilingual Man-Computer Interfaces and Computer-Assisted Communication areas; Automatic Language Identification is consequently a challenge for the user-friendly society of the XXIst century.

The state of the art approach is based on phonotactics modeling. Nowadays, best systems identify quite well (90% of correct identification) one language among eleven, with 45s duration records [5,6,7]. Even if the results are good, only marginal improvements have been performed since '96, and it is now crucial to consider alternative features to overcome these limits [8]. Especially, it is necessary to reduce the record duration, and to reach these purposes, the classical ALI system must be improved, and the wide range of distinctive features available to characterize a language (phonetic [5], phonotactic [9], prosodic [10], lexical, morpho-syntactic features) must certainly be exploited.

European researches on this topic are still rare [11,12,8], while Europe is deeply a multilingual area. A few laboratories are already involved in ALI projects across the Europe, but the SPLC SIG is the ideal framework for the rise of a world network consisting of the different researchers (linguists, neuroscientists, speech processing engineers, etc.).

### 3.4. Commercial Applications for Speaker and Language Recognition

Recently, an increasing interest in commercial applications of techniques such as speaker verification, speaker identification and language recognition can be observed. In a world of mobile, networked users, these techniques offer novel solutions for increased security and user friendliness.

SpLC can help in the process of bringing these techniques into products by:

- Enhancing the communication between researchers and technology providers;
- Giving an overview on ongoing research efforts as well as existing prototype implementations in these areas;
- Providing a link to ongoing standardization organizations and projects (e.g. COST).

In general, the described technologies are only part of products. Successful integration of speaker verification requires a thorough understanding of the technical possibilities and limitations as well as the application.

Speaker verification inherently can not provide 100% accuracy. Therefore the combination with other security techniques, in particular other biometric methods, is crucial. Furthermore, such novel methods raise a number of legal questions. SpLC will provide a forum to discuss these interdisciplinary questions by bringing together experts from the various fields. The discussion during the workshop in RLA2C workshop in 1998 was an important step in this direction. The workshop, however, also showed clearly that further discussions are needed to develop a common language and understanding between the various fields.

## 4. Transversal/special activities

### 4.1. Evaluation group

The aim of the evaluation group is to work on the evaluation paradigm, for all the various axes of the SpLC domains. Particularly, the group wants to highlight evaluation tasks for commercial applications, for speaker tracking and indexing, and for language identification.

For each task, the evaluation group intends to propose an evaluation scheme, a format for the representation of the results, and when possible some programs to do the scoring.

Especially for language identification, the task will be perhaps more difficult because there is no precedent, and the evaluation scheme cannot be an improvement of any previous one; databases will be one of the key point [13].

The evaluation group also intends to work if possible in collaboration with other evaluation projects (like NIST/NSA). A web page focused on these issues will be available very soon on the web site of SpLC.

### 4.2. Meetings and Workshops group

As said in the introduction, SpLC SIG was launched during the RLA2C workshop, in 1998. This point illustrates the necessity of thematic workshops to define the state-of-the-art and to increase collaboration and exchanges. Workshops as RLA2C (Avignon, 98) or Martigny ETRW (94) are two good examples. The main role of SpLC is to organize next workshops with the objectives to guaranty the independence of the workshops and the continuity of the editorial policy.

The next workshop, the first SpLC event (2001: A Speaker Odyssey), takes place in Crete (Greece) during June 2001 (18-22). Extended information is available at <http://www.odyssey.westhost.com/>.

During this workshop, SpLC also organizes its first plenary meeting.

### 4.3. Speaker Indexing

Speaker indexing consists in extracting information about speakers from audio documents. This is still a novel task in speaker recognition. The first papers on the topic were presented only about 5-10 years ago [14,15,16]. It has been proposed by NIST for the speaker recognition evaluation workshop only 2 years ago. And this topic is regularly presented in conferences only for 1 or 2 years now.

New algorithms need to be developed, evaluation procedures need to be defined, and databases for evaluation need to be designed, built, and made available. The speaker indexing group of SpLC proposes to be a place to discuss all these issues.

## 5. SpLC Web Site

A new web site has been designed in order to facilitate the circulation of information between SpLC members. The web site is composed of several pages. The Home page lists the four main areas of SpLC, the various specialized groups, and the recent news of SpLC. The other pages of the site include a list of events organized by SpLC, the list of the board members, a page to subscribe automatically to SpLC, the constitution of SpLC, a way to contact the SpLC board, a few interesting links including some links to related workshops and conferences or to laboratories working on speaker and language characterization, and a bibliography of the domain with the possibility to add new references.

The web site is currently hosted by the Laboratoire Informatique d'Avignon at <http://www.lia.univ-avignon.fr/heberges/SPLC>. The site can also be reached from the web page of ISCA dedicated to the SIGs at <http://www.isca-speech.org/sig.html>. Questions and suggestions on the web site are very welcomed and can be sent directly to the webmaster of the site. Contributions from the specialized groups and from all the members of SpLC are also waited.

## 6. Conclusion

During the first year of SpLC SIG, a lot of work has already been accomplished. A workshop has been organized (2001: A Speaker Odyssey), a web site has been designed and is going to present regularly the work done in the framework of SpLC, and several discussion groups have been initiated.

However, there are still a lot of things to do. Every main area should have its own leader and its own electronic

means, as for the specialized groups. New propositions need also to be done in order to keep SpLC very active in the domain and up-to-date. If you are interested in speaker and language characterization, do not hesitate to visit the web site, to subscribe, or to send your comments or questions. SpLC faces a very interesting challenge: be the place for stimulating exchange and innovative research thanks to the contribution and participation of all its members.

## 7. Reference

- [1] J. Luetttin, N.A. Thacker, and S.W. Beet. Speaker identification by lipreading. *Proceedings of ICSLP 96*, vol. 1, pp. 62-65, 1996.
- [2] Claude C. Chibelushi, Farzin Deravi, and John S.D. Mason. A review of speech-based bimodal recognition. Accepted for publication in *IEEE Transactions on Multimedia*.
- [3] J.D.Brand, J.S.Mason, and S.Colomb. Visual speech: a physiological or behavioural biometric? *Proceedings of AVBPA 2001*, Halmstad, Sweden, 2001.
- [4] G. Roethenbaugh. "Biometrics Explained", <http://www.icsa.net/services/consortia/cbdc/explained.shtml>, 1999.
- [5] T. J. Hazen and V. W. Zue. Segment-based automatic language identification. *Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2323-2331, 1997.
- [6] J. Hieronymus and S. Kadambe. Robust spoken language identification using large vocabulary speech recognition. *Proceedings of ICASSP '97*, vol. 2, pp. 1111-1114, Munich, Germany, 1997.
- [7] M. A. Zissman. Comparison of four approaches to automatic language identification of telephone speech. *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 1, Jan 1996.
- [8] Michel Dutat, Ivan Magrin-Chagnolleau, and Frédéric Bimbot. Language recognition using time-frequency principal component analysis and acoustic modeling. *Proceedings of ICSLP 2000*, Beijing, China, October 2000.
- [9] H. Kwan and K. Hirose. Recognized phoneme-based N-gram modeling in automatic language identification. *Proceedings of EUROSPEECH '95*, pp. 1367-1370, Madrid, Spain, September 1995
- [10] S. Itahashi and L. Du. Language identification based on speech fundamental frequency. *Proceedings of EUROSPEECH '95*, pp. 1359-1362, Madrid, Spain, September 1995.
- [11] L. Lamel and J.L. Gauvain. Language identification using phone-based acoustic likelihood. *Proceedings of ICASSP '94*, pp. 293-296, Adelaide, Australia, 1994.
- [12] F. Pellegrino and R. André-Obrecht. From vocalic detection to automatic emergence of vowel systems. *Proceedings of ICASSP '97*, Munich, Germany, April 1997
- [13] T. L. Lander et al. The OGI 22 language telephone speech corpus. *Proceedings of EUROSPEECH '95*, pp. 817-820, Madrid, Spain, September 1995.
- [14] Man-Hung Siu, George Yu, and Herbert Gish. An unsupervised, sequential learning algorithm for the segmentation of speech waveforms with multiple speakers. *Proceedings of ICASSP '92*, vol. 2, pp. 189-192, San Francisco, USA, March 1992.
- [15] Lynn Wilcox, Francine Chen, Don Kimber, and Vijay Balasubramanian. Segmentation of speech using speaker identification. *Proceedings of ICASSP '94*, pp. I.161-I.164, 1994.
- [16] Aaron E. Rosenberg, Ivan Magrin-Chagnolleau, S.Parthasarathy, and Qian Huang. Speaker detection in broadcast speech databases. *Proceedings of ICSLP '98*, Sydney, Australia, December 1998.