Polyphone-IPSC-01: a shared speakers database for evaluation of forensic automatic speaker recognition systems

Didier Meuwly*, Anil Alexander**, Andrzej Drygajlo** and Filippo Botti***

* The Forensic Science Service, Trident Court, Solihull Parkway, Solihull B37 7YN, UK

** Institute of Signal Processing (ISP), Swiss Institute of Technology (EPFL), 1015 Lausanne, Switzerland

*** Institut de Police Scientifique et Criminologie (IPSC), BCH, University of Lausanne, 1015 Lausanne, Switzerland

Abstract

This paper describes the content and the use of the speaker database Polyphone-IPSC-01. It also proposes to share this database for research purposes. Decision and terms of agreement to share the database remains a decision of the owner of the database, the School of Forensic Science of the University of Lausanne.

Keywords

speaker recognition, validation, database

Introduction

Forensic automatic speaker recognition systems can be used for intelligence and evidence purposes. However their performance need to be evaluated before the systems can be used in operations. Following [1], the validation process is there are 3 stages process constituted of technology, scenario and operational evaluation:

1. The goal of a technology evaluation is to test the system to determine the performance of the algorithms in ideal conditions
2. The goal of scenario evaluation is to test the system in relevant scenarios (depending on the task) to determine the performance of the system in simulated operational conditions
3. The goal of operational evaluation is to determine the performance of the system in operational conditions

The evaluation protocols used for the validation will depend on the task, intelligence or evidence, but we consider as good practice to design them following the criteria defined in the the Daubert decision [1], e. g. the falsifiability criterion [2-3].

Evaluation protocol to test the evidence task

We propose a protocol to test forensic scenarios as part of the validation process of forensic automatic speaker recognition systems used to produce evidence. This protocol is based on the calculation of likelihood ratio of speech evidence regarding two hypotheses: the prosecution hypothesis (H_p) and the defence hypothesis (H_d). Two experiments are conducted: in the first experiment H_p is true, what means that the suspected speaker is the source of the trace submitted for analysis and in the second experiment H_d is true, what means that the source of the trace submitted for analysis is another speaker (see Figure 1). The results obtained for the two experiments are plotted in form of Tippett plots or in form of cumulative density functions (see Figure 2).

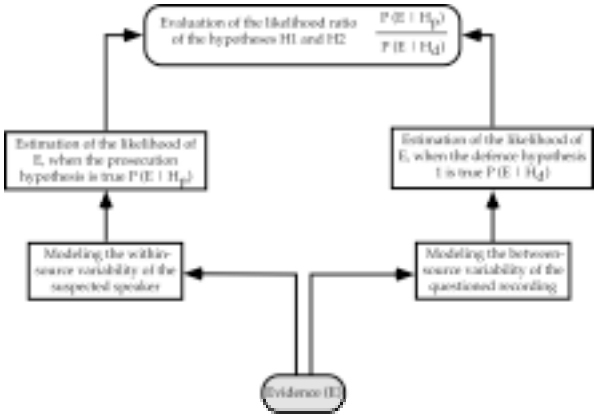

Figure 1: Design of the evaluation protocol

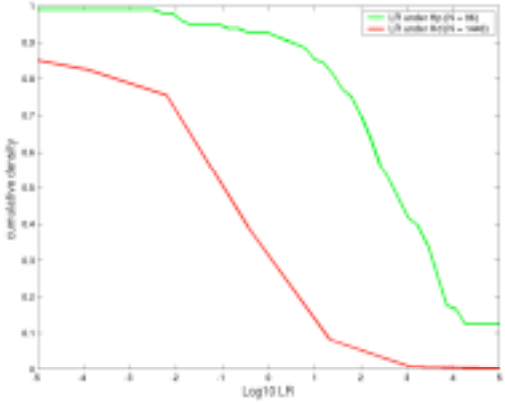

Figure 2: Tippet plot representation, obtained using the Polyphone-IPSC-01 database. Conditions: 96 tests (16 male speakers * 6 tests), spontaneous speech, PSTN

Data sets requested to test the evidence task

3 + 1 data sets are necessary to run these two experiments described above (see Figure 3). the 3 datasets normally requested for the operational use of forensic automatic speaker recognition systems (1, 2 and 3) and the and one data set only requested for evaluation (4).

[1] Daubert vs Merell Dow Pharmaceuticals, Inc.," 509 U. S. 579, 1993

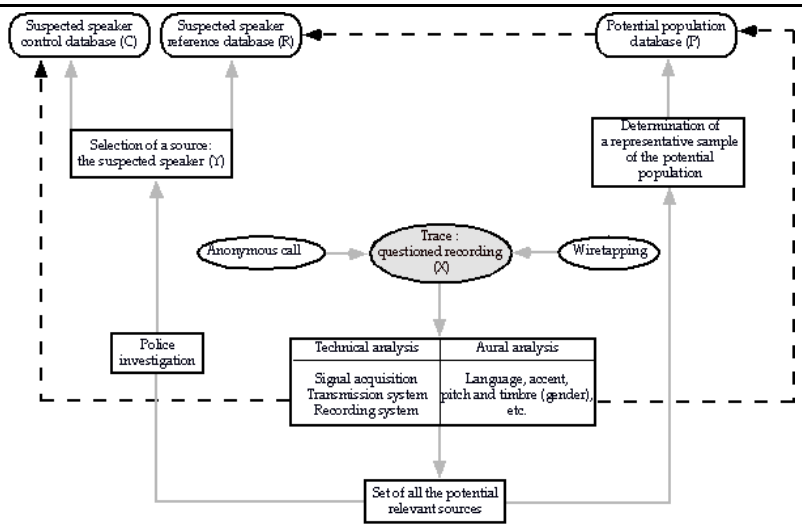

Figure 3: Data sets requested for a scenarios evaluation of the evidence task:

1. **The potential population data set (P),** used to model the between-sources variability of all the potential sources of the trace,
2. **The suspected speaker reference data set (R)**, used to model the speech of the suspected source of the trace,
3. **The suspected speaker control data set (C)**, used to evaluate the within-source variability of the suspected source of the trace,
4. **The forensic traces data set (X)**, used to reproduce traces recorded in forensic operations

Content of the Polyphone-IPSC-01 database

The potential population data set (P)

As the recording of a large-scale speech data set is a long and expensive procedure, we have decided not to record the P data set, but to select an existing one. We have chosen the Swiss-French Polyphone database of Swisscom©: 4500 speakers (2500 females and 2000 males) have recorded one session of read and spontaneous speech (80-120s) through the public switched telephone network (PSTN). This database is available through The European Language Resources Association (ELRA).

The 3 other data sets (R, C and X)

The 3 other data sets (R, C and X), constituting the "Polyphone-IPSC-01 database" have been recorded during [4] with 32 Swiss-French speakers (8 couples of "sound-alike" females and 8 couples of "sound-alike" males

Information about the speakers

Female couples (A – H)

Couple	Speaker	Mother language	Age	Partner	Relationship	Subjective judgment of the auditive closeness
A	F00	German	24	F01	Daughter	Greater closeness
	F01	German	54	F00	Mother	in German than in French
B	F04	French	32	F06	Daughter	Great proximity
	F06	French	59	F04	Mother	through the telephone
C	F05	French	26	F49	Sister	Great proximity
	F49	French	27	F05	Sister	through the telephone
D	F07	French	31	F08	Sister	Great proximity
	F08	French	33	F07	Sister	through the telephone
E	F09	French	64	F33	Mother	Reasonable proximity
	F33	French	32	F09	Daughter	through the telephone
F	F32	French	13	F44	Twin sister	Great proximity
	F44	French	13	F32	Twin sister	through the telephone
G	F54	French	52	F55	Mother	Reasonable proximity
	F55	French	26	F54	Daughter	through the telephone
H	F58	French	54	F59	Daughter	Reasonable proximity
	F59	French	25	F58	Mother	through the telephone

Table 1: the 8 couples of "sound-alike" female speakers

Male couples (I – P)

Couple	Speaker	Mother language	Age	Partner	Relationship	Subjective judgment of the auditive closeness
I	10	French	61	56	Father	Great proximity
	56	French	29	10	Son	through the telephone
J	11	French	31	20	Son	Great proximity
	20	French	62	11	Father	through the telephone
K	12	French	30	40	Son	Reasonable proximity
	40	French	55	12	Father	through the telephone
L	13	French	36	41	Brother	Great proximity
	41	French	40	13	Brother	through the telephone
M	14	French	33	15	Brother	Great proximity
	15	French	35	14	Brother	through the telephone
N	16	French	23	17	Son	Great proximity
	17	French	56	16	Father	through the telephone
O	18	French	33	19	Twin brother	Great proximity
	19	French	33	18	Twin brother	through the telephone
P	22	French	77	39	Father	Reasonable proximity
	39	French	49	22	Son	through the telephone

Table 2: the 8 couples of "sound-alike" male speakers

Information about the speech samples

1. The R database is constituted of 7 reference samples per speaker (80–120s of read and spontaneous speech) recorded in a period of 1 to 3 months. 6 samples are recorded through the PSTN and 1 through the GSM network
2. The C database is constituted of 13 to 47 samples of spontaneous speech (1–28s) depending on the speaker. These samples are recorded through the PSTN in one session
3. The X database is constituted of traces reproducing realistic forensic conditions.

   Each of the 32 speakers has recorded 6 spontaneous speech samples (5 with PSTN and 1 with GSM), 1 anonymous call without disguise (PSTN) and 1 with free disguise (PSTN). Finally, the first spontaneous speech sample recorded through the PSTN has been used to produce a sample recorded with an analogue recording machine from the police and to produce 8 noisy samples with respectively: 30, 24, 18, 12, 9, 6, 3 and 0 dB of signal-to-noise ratio.

Conclusion

This database can be used to evaluate the performance of automatic speaker recognition systems under forensic scenarios. We propose to share this database to allow the comparison of different systems using the same data. Formal requests for the sharing should be addressed with project details to the School of Forensic Science of the University of Lausanne, owner of the database. Decision and terms of agreement to share the database remains a decision of the owner of the database.

Bibliography

1. **MANSFIELD A. J., WAYMAN J. L.** (2002) « *Best Practices in Testing and Reporting Performance of Biometric Devices* », Issue 2 draft 9, Issued 08/02, United Kingdom Government Biometric Working Group. http://www.cesg.gov.uk/technology/biometrics

2. **LOEVINGER, L.** (1995) « *Science as evidence* », Jurimetrics, Journal of law, science and technology, vol. 35, no. 2, pp. 153 - 190.

3. **BLACK, B., AYALA, F. J., SAFFRAN-BRINKS, C.** (1994) « *Science and the law in the wake of Daubert: A new search for scientific knowledge* », Texas Law Review, vol. 72, no. 4, pp. 715 - 802.

4. **MEUWLY D**; (2001) « *Reconnaissance de locuteurs en sciences forensiques: l'apport d'une approche automatique* », Ph D thesis, University of Lausanne.