DATABASES FOR SPEAKER RECOGNITION: ACTIVITIES IN COST250 WORKING GROUP 2

Håkan Melin

KTH, Centre for Speech Technology (CTT),
Drottning Kristinas väg 31, SE-100 44 Stockholm, Sweden
Email: melin@speech.kth.se

ABSTRACT

Working Group (WG) 2 of the COST250 Action "Speaker Recognition in Telephony" has dealt with databases for speaker recognition. The present paper gives an overview of the activities in this WG, and presents its main results. The first result is an overview of 36 existing databases that has been used in speaker recognition research. Those include both public and proprietary databases. As part of the overview, some of the variability represented in those databases is analyzed. The second result is the publicly available Polycost database, a telephony-speech multi-session database with 134 speakers from all around Europe. Together with pre-defined experiment specifications, this database is a useful resource to aid in the assessment of speaker recognition systems in general, and in comparing systems across sites, in particular.

1. INTRODUCTION

The availability of good speech databases is crucial for development and assessment in all branches of speech research. In the beginning of the COST250 Action, in January 1995, Working Group 2 (WG2) was formed to deal with questions related to the availability, design, collection and production of speech databases for speaker recognition. The present paper gives an overview of the activities and results of WG2.

The three main activities were first to list and characterize speech databases that were available to COST250 participants; then to design, record and produce the now publicly available Polycost database; and finally, to do case studies of several database projects. Section 2 of this paper gives an overview of many of the existing databases that has been used in speaker recognition research, while section 3 presents a basis for detailed characterization of speaker recognition databases. The characteristics of the Polycost database is then given in section 4, along with an outline of the various stages traversed to create this database. A list of publications where Polycost has been used in research is also given. Section 5 summarizes the case study presentations made during WG2 technical sessions. The Working Group's contribution to the research field of speaker recognition is finally summarized in section 6.

2. DATABASE SURVEY

The first task set out for WG2 was to survey what databases were available to each COST250 partner; what was their characteristics, and was there a database that could potentially be used by all partners? The first step was to define a basis for the survey in terms of a series of questions, and compile them into a questionnaire. Secondly, the questionnaire was distributed among partners, and finally, answers to the questionnaire was collected and compiled into a comprehensive report. This report covers 18 databases and was finalized in June 1996 (Lindberg et al., 1996).

Table 2 contains a comprehensive (not necessarily exhaustive) list of speech databases that have been used in speaker recognition research around the world, with an emphasis on data from European laboratories. In many, not all, cases the databases are available to other researchers. The list includes databases from the above mentioned survey, but also several other databases that were either missing from the survey or have been produced after the survey was completed. For databases that were developed further since June 1996, information in the table has been updated with respect to the original survey.

It is interesting to look at the variability the various databases in Table 2 represent. Most of the databases are multi-session databases that capture temporal intra-speaker variability. Several contain recordings from the same speaker during three months or more (often motivated by an investigation in (Furui, 1986) where it is reported that the size of a speaker's feature subspace increases during the first three months of measurement and is fairly constant after that). In some cases, speakers have been recorded during more than a year: 1½ years in PolyVar and Gandalf, and 2 years in the CSLU Speaker Recognition Corpus (references to database publications are collected in Table 2).

Many multi-session databases contain telephone handset variability in addition to temporal intra-speaker variability. In Ahumada, for example, speakers were recorded three times through dual microphones in a studio, and three times over telephone lines from different locations and handsets each time. In Gandalf, speakers placed half the calls from a single "favorite" handset and half the calls from 2-9 other handsets during the first four months of the recording period. With this design, temporal intra-speaker variability and handset variability can be studied in parallel. In the Switchboard-II, 'LIMSI/CNET', CSLU and SpeechDat speaker recognition databases, special attention has also been paid to handset variation. Handset variability is often accompanied with variability in caller environment.

A single-session database can be useful when studying a particular (non-temporal) variability in isolation. A few of the databases in Table 2 have been designed to include such variability; two with intra-speaker variability and several with variability in recording conditions:

- VeriVox: voluntary intra-speaker variation from weak, strong, slow, fast, and denasalised speech,

and involuntary variation from speech in noise and speech under stress (Karlsson et al., 1998),
- ‘Disguised voices’: the effect of nine different techniques for voice disguise,
- LLHDB, HTIMIT and SR4X: telephone handset variation,
- TSID: radio transmitter and receiver variation,
- NTIMIT and CTIMIT: telephone circuit variation (in fixed and mobile networks respectively).

Note that a single-session database is not useful for estimating the absolute level of performance for a speaker recognition system in a practical application because it will never include normal temporal intra-speaker variability.

Two of the databases in Table 2 contain speech from the same speaker in two languages. The Ahumada-Mataro database contains speech from bilingual speakers of Castilian Spanish and Catalan, and speakers in Polycost speak both English and their mother tongue.

Regarding inter-speaker variability, it is of course desirable to include many speakers to achieve a good sampling of a speaker population. At the same time, it is often desirable to get a good sampling of individual speakers over multiple sessions to study intra-speaker variability. With only limited resources for database creation, a trade-off between the number of speakers, number of sessions and total cost is often necessary. A common design technique, used in many databases (SpeechDat, Switchboard-1, PolyVar, ‘LIMSI/CNET’, SIVA, Broertjes-Polyphone, and Gandalf), is to record a smaller set of speakers in many sessions and a separate, larger set of speakers in a single session. With this technique, one can achieve both good sampling of inter-speaker variability in potential impostor speakers, and of intra-speaker variability in client speakers. Note that the same effect may be achieved if a multi-session speaker recognition database is used together with a database designed for speech recognition purposes. This is the case with, for instance, the SpeechDat and Broertjes-Polyphone databases.

Other interesting aspects of inter-speaker variability is the inclusion of close relatives among speakers, and of human or technical mimicry. In the Broertjes, Brent, Gandalf and SpeechDat databases, pairs of close relatives, such as twins, siblings, father-son and mother-daughter, have been included. Mimicry has been included in the CSLU database, where in each call a speaker is asked to imitate a given prompt phrase. In the literature are also studies where speech from one person has been transformed by technical means to impersonate another speaker. In for instance (Lindberg and Blomberg, 1999) two speakers from Gandalf were used as target speakers, and in (Genoud and Chollet, 1999) 18 target speakers from PolyVar were used.

Recent work by (Campbell Jr. and Reynolds, 1999) gives a good and more detailed overview of currently publicly available databases for speaker recognition evaluation. Another database overview is (Godfrey et al., 1994).

3. DATABASE CHARACTERIZATION

An important secondary outcome of work with the database survey is a series of questions for characterizing a speaker recognition corpus. The questionnaire was updated during 1998 with several new questions, and the complete updated questionnaire is included in (Melin, 1999). Questions are divided into the following topics: 1) name and availability, 2) general information, 3) speaker material (including questions on the number of speakers, inter-speaker variation, intra-speaker variation, and impostor characterization), 4) speech contents, 5) post-processing, 6) recording equipment, 7) recording environment, 8) publications, and 9) other information. An example of a corpus description based on the questionnaire is included in section 4.3, where the Polycost database is described.

Other work on the characterization of databases was presented in (Falcone, 1995b). A strategy for how to characterize speaker recognition databases was presented, including measures of acoustic properties over time and frequency. Special attention was dedicated to possible ad hoc measures using results from a standardized speaker recognition system (a reference system). More information is available in (Falcone and Contino, 1995). The ideas around the use of a reference system have been further developed in WG4 (Falcone et al., 1999).

4. POLYCOST

A conclusion from the database survey was that none of the existing databases was both 1) suitable as a common research corpus within COST250, and 2) could be made available to everybody. It was decided that some effort should be dedicated to creating a common database for the COST250 partners. It was argued that if all partners had access to a common database, it would increase the possibilities for comparing work done by different partners. A database together with standardized experiment specifications for it would allow collaborative research to a greater extent than if each partner worked with its own data, protocols and algorithms. Such a database would therefore be a good support for work in the other working groups. Finally, creating the database would be a practical exercise that should give useful insights into many of the problems that the Working Group was set up to deal with. To this end the Polycost database (Petrovska et al., 1998) (Hennebert et al., 1999) (Campbell Jr. and Reynolds, 1999) was subsequently created.

The following sections describe each of the steps in creating the Polycost database. They also give a short summary of the properties of the database and cite several references where it has been used in experiments. The latest information on Polycost is published on the Polycost home page at http://circwww.epfl.ch/polycost.

4.1 Design, recording and distribution

During 1996 the Polycost database was designed, recorded and distributed among COST250 members as a first preliminary CD-ROM version. The design was based on the Polycode database previously recorded at IDIAP (Switzerland). Recording was done in Switzerland at IDIAP and at the Signal Processing Laboratory (LTS) at EPFL. The recording platform developed at IDIAP was based on a Sun XTL platform and recorded data off an ISDN subscriber line. Speakers were recruited from each of the participating countries. The first preliminary version was produced at KTH (Sweden) with sponsoring from Telia InfoMedia AB (Sweden). For the final distribution of Polycost a contract has been made with the European Language Resources Association (ELRA). The database will be distributed for commercial and research purposes to a wide audience at a low cost. Part of the revenue from distribution is used to finance the annotation work at Circuits and Systems Laboratory (CIRC) at EPFL.

4.2 Post processing

The annotation work was done during 1997 to 1999 (Petrovska et al., 1996)(Hennebert, 1997)(Hennebert el al., 1998). The mother tongue utterances were transcribed by several COST250 partners. The utterances spoken in English were annotated at CIRC/EPFL and KTH. The initial goal was to provide word-level transcriptions together with word segmentation information, all to be produced at EPFL (Petrovska et al., 1996). This goal was later revised to include only transcriptions; producing the segmentation information required more manual labor than expected, and would take too long time to complete, given the resources available at EPFL. Segmentation information is thus provided for only a subset of the files in the database while a transcription is available for every file.

4.3 Description

This section gives a condensed description of Polycost based on the questionnaire for characterizing speaker recognition corpora referred to in section 3.

1. **Name and availability:** Polycost was recorded in Switzerland during January-March 1996. A first release on two CD-ROM was made in June 1996 within COST250, and has been publicly available from ELRA at a low price since 1998. The next release, including annotations, is planned for the end of 1999.
2. **General information:** the corpus contains speech data and was designed for general-purpose experiments in speaker recognition. It is also suitable for language and accent identification and speech recognition experiments.
3. **Speaker material:** 74 male and 60 female speakers. **Inter-speaker variation:** Most of the speakers are actively involved in speech research. 85% were between 20 and 39 years old. Hence, the speaker group exhibits relatively small variation in age, profession and educational background. On the other hand, variation in language and accent background is large. A summary judgement of inter-speaker variation on a scale 1-5 would therefore be around 3. **Intra-speaker variation:** Three speakers were recorded in 2-5 sessions, 105 speakers in 6-10 sessions, and 25 speakers in 11-15 sessions. Most speakers were recorded during 2-3 months. The effective duration of speech in each call is approximately 60 seconds. **Impostor characterization:** casual impostors only.
4. **Speech contents, language:** All speakers are recorded both in English and in their mother tongue. 85% are non-native speakers of English. 17 languages are represented: nine with five speakers or more (French, English, Dutch, Turkish, Swedish, Italian, Danish, Spanish and Portuguese) and eight with less than five speakers (Catalan, Arabic, Russian, Polish, Macedonian, Lithuanian, Galician and German). **Text material:** Out of 14 files in each session, 12 are text-dependent and two are text-independent. Among the text-dependent files are 10 sequences of digits (including one telephone number) and two sentences. All are fixed from session to session and are spoken in English. Text-independent items are spoken in the speaker's mother tongue: one in response to a written question and the other as a free monologue.
5. **Post-processing:** All files have an orthographic transcription. Six of the digit sequences also have a verified word-level segmentation.
6. **Recording equipment:** All sessions are recorded from telephone calls made from unknown types of handsets to a Euro-ISDN digital subscriber line. Around 80% of the speakers made all their calls from the same handset.
7. **Recording environment:** most calls were made from a home or office environment with varying (uncontrolled) levels of noise.
8. **Publications:** [63] and [31] describe the database, while [46] and [51] define baseline experiments. Table 1 lists references with recognition results.
9. **Other information:** Special features are the use of international telephone lines, non-native speakers of English, and speech in both English and the speaker's mother tongue.

4.4 Exploitation

A set of four standard baseline experiments has been defined for Polycost. These were first presented as a version 1.0 in (Melin and Lindberg, 1996). Following initial results (Nordström et al., 1998), the specification was subsequently revised to a version 2.0 (Melin and Lindberg, 1999). The specification of the baseline experiments is further treated within the scope of WG4 (Falcone et al., 1999).

Results from experiments on Polycost have been reported at COST meetings (Olsen and Lindberg, 1999) and at major international conferences. An overview of those reports is given in Table 1.

5. CASE STUDIES

During the course of the COST 250 Action, several presentations were made on the design, recording procedure, and post-processing techniques used in several speaker recognition database projects. Those presentations were quite suitable as case studies for the Working Group. A summary

BE	**Task description**	**Ver. 1.0**	**Ver. 2.0**
1	TD verification, sentence	[32][33][55][56]	[56]
2	TD verification, 10 digits	[47][48][55][56]	[49][56]
3	TI verification	[15][55][56]	[56]
4*	TI identification	[1][2][3][12][13][14][28][42]	

Table 1. An overview of publications that report on results from baseline experiments (BE) 1-4 with Polycost. 'TD' indicates a text-dependent and 'TI' a text-independent task. *Note that experiments on speaker identification (last row) are made with variations of BE4, with the main deviation from BE4 specification being the choice of different target speaker subsets.

of each presentation is given in this section, along with references to more information on the various databases.

SIVA: The SIVA Italian telephone-speech database was designed for experiments on speaker recognition. It contains as many as 691 speakers recorded in 1-26 sessions each. The design of the database, the hardware and software setup of the recording system, and the definition and realization of a pilot experiment were described in this presentation (Falcone, 1995a). Further information about SIVA is available in (Falcone and Gallo, 1996).

Gandalf: The Gandalf telephone speech database is designed for experiments on speaker recognition with special attention to intra-speaker and handset variability. It contains 86 client speakers recorded in 17-29 sessions over a period of 1½ years, and 83 impostor speakers. Besides audio files, a lot of information related to speakers and sessions have been collected, such as the type of telephone handset, a characterization of background noise, and occurrences of head colds and sore throats. In this presentation (Melin, 1995), the database design and experiences from the first stage of data collection was communicated. More information on this database is available in (Melin, 1996).

SpeechDat: SpeechDat (II) is an EU-funded project (LE2-4001) with the goal of producing large telephone speech databases in several European languages (Höge et al., 1997). Three categories of databases are recorded: 5000 speaker fixed telephone network, 1000 speaker mobile telephone network and 120-speaker/20-session speaker verification database. This presentation (Lindberg, 1996) described the design and specification of the speaker verification databases. More information is available in (Nataf, 1996).

M2VTS: In part of a presentation (García-Plaza and Fernández, 1996) from a representative of the M2VTS project (ACTS, AC102), the initial efforts of creating a multi-modal database for speaker recognition was described. The database contains the speech signal plus video sequences from three viewing angles. At the time of this presentation, 37 French subjects had been recorded, but up to 300 subjects was planned for. More information on this database is available in (Pigeon and Vandendorpe, 1997). An extension database, XM2VTS, has later been created with 300 British English speakers (Messer et al., 1999).

Polycost: The annotation method used for Polycost digit utterances was presented (Petrovska et al., 1996). With this method, only 16% of the seven-digit utterances and 30% of the ten-digit utterances had to be manually annotated. Annotation in this case involves both transcription and marking word boundaries in time. In summary, a connected digit recognizer was iteratively trained on parts of the data. This recognizer was then applied to all digit utterances, and for those utterances where the recognized sequence was identical to the manuscript sequence, the recognition result was taken as the correct annotation for the utterance. Results for the remaining utterances were used as templates for manual annotation.

SESP is a Dutch telephone-speech database designed for experiments on speaker recognition. It contains speech from 45 speakers recorded in 21-32 sessions each. Each speaker placed calls from a variety of handsets and from many types of locations. A substantial proportion of the calls came from foreign countries. During the presentation (Kuitert, 1996), the database was described and some preliminary results from the CAVE project were given. SESP was the main research corpus during the second half of the CAVE project (LE1-1930) (Bimbot et al., 1998). SESP and an extension, SESP 2, are used extensively in the PICASSO project (LE4-8369) (Bimbot et al., 1999).

AHUMADA is a Castilian Spanish speaker recognition corpus designated to commercial and forensic tasks. It was designed with special attention to intra-speaker variability and external variability, and currently contains 104 male and 80 female speakers recorded in six different sessions. The design and collection of the database was described (Ortega-García, 1999), and results were presented for a text-independent verification task with a 25-speaker subset and a GMM-based recognition system. Influences on the error rate from changes in speaking-style and microphone type were demonstrated. An expansion of the database is currently underway. This database is further described in (Ortega-García et al., 1998).

6. CONCLUSION

The COST250 Action has been an important forum for the discussion of ideas on the design and implementation of databases for speaker recognition. It has given its participants an insight into several problems involved in the creation of such databases, and into various approaches to solving those problems. The Action has also provided its participants and the wider research community with a good overview of several existing, public and proprietary, databases. Last, but not least, the Action has provided the research community with Polycost, a new public database created for research in speaker recognition. It is our hope and belief that this database will be a useful speech resource that will contribute to the advances of state-of-the-art in speaker recognition long after the end of the COST250 Action.

7. ACKNOWLEDGEMENT

Fred Lundin was coordinating WG2 from 1995 to 1997, during which time the strategy for the Working Group was drawn and most of the work was planned and initiated. The author acted as coordinator 1998-1999. The coordinators wish to thank everybody who contributed to this Working Group. This includes among many others: those who gave their voice to Polycost; those who submitted information to the database survey and those who compiled it; students who worked on Polycost annotation; those who transcribed mother tongue utterances in Polycost; presenters during technical meetings; and, most of all, Dijana Petrovska, Jean Hennebert and Dominique Genoud for all their work with Polycost.

Name	Creator(s)	Distributor	Language	Cat. 1 male	Cat. 1 female	Cat. 2 male	Cat. 2 female	Recording	Digits	Words	Sentences	Spontaneous	References	96 survey
EUROM-1, Danish	Tele Danmark, CPK	ELRA	da	5	5	30	20	mic	•		•		[40]	•
Polycost	COST250	ELRA	en + 16	74	60	-	-	tel	•		•	•	[31,51]	
Brent	BT		en GB	50	50	-	-	tel	•	•	•		[5,61]	•
Millar	BT		en GB	63		-	-	mic	•				[4]	
SpeechDat (FDB+SDB)	GPT Limited	ELRA	en GB	60	60	2500	2500	tel	•	•	•		[34,54]	
XM2VTS	Univ. of Surrey, M2VTS-project		en GB	295		-	-	mic, video	•				[6,53]	
CSLU Speaker Recognition Corpus	OGI/CSLU	OGI	en US	1) 47	1) 53	-	-	tel	•	•	•	•	[11]	
Disguised voices	Hollinell, Meverly		en US	-	-	7	5	tel, mic			•		[40]	•
KING-92	ITT	LDC	en US	51	0	-	-	tel, mic				•	[10,27]	
LLHDB	MIT-LL	LDC	en US	-	-	24	29	mic 3)			•	•	[65]	
SR4X	OGI/CSLU	OGI	en US	-	-	36		tel		•			[57]	
Switchboard-1 (incl. SPIDRE subset)	Texas Instruments, NIST, LDC	LDC	en US	2) 22	2) 23	2) 280	2) 218	tel				•	[10,26]	
Switchboard-2, phase I	LDC	LDC	en US	358	299	-	-	tel				•	[10]	
Switchboard-2, phase II	LDC	LDC	en US	679		-	-	tel				•	[10]	
TIMIT (+N/C/H/FFM-TIMIT)	MIT, SRI, TI (+ others)	LDC	en US	-	-	630		mic (+tel)			•		[10,65]	
TSID	MIT-LL, LDC	LDC	en US	-	-	31	4	mic, radio	•		•		[38]	
YOHO	ITT, Oklahoma State Univ.	LDC	en US	106	32	-	-	mic	•				[10]	•
Ahumada	Univ. Politéc. Madrid		es	1) 104	1) 80	-	-	tel, mic	•	•	•	•	[59]	
Ahumada-Mataro	Univ. Politéc. Mataro		es, ca	1) 49		-	-	tel, mic	•	•	•	•	[68]	
TelVoice	Univ. of Vigo		es	39	20	-	-	tel	•		•	•	[66,67]	
M2VTS	UCL, M2VTS-project	ELRA	fr BE	25	12	-	-	mic, video	•				[6,64]	
LoCoMic	IDIAP		fr CH	22		-	-	mic	•		•	•	[6]	
Polycode	IDIAP		fr CH	10	10	-	-	tel	•					•
PolyVar (incl. SpeechDat subset)	IDIAP	ELRA	fr CH	43	28	42	30	tel	•	•	•		[10,43]	•
'LIMSI/CNET'	CNET, LIMSI, Vecsys		fr FR	100		1000		tel	•	•	•	•	[24,37]	
RECLOC	CRIN/INRIA		fr FR	17	7	-	-	mic			•		[41,52]	•
SpeechDat (FDB+SDB)	Matra Communication	ELRA	fr FR	60	60	2500	2500	tel	•	•	•		[34,54]	
SUBTV 4)	CRIN/INRIA		fr FR	-	-	161	63	mic	•	•			[40]	•
SIVA	FUB	ELRA	it	207	229	128	127	tel	•	•	•		[10,20]	•
Broertjes+Polyphone 5)	KUN+KPN Research		nl	100	0	2616	2434	tel	•	•	•		[9]	•
SESP	KPN Research		nl	23	22	1	0	tel	•		•		[7,9]	•
SESP 2	KPN Research		nl	84	64	-	-	tel	•				[8]	
SESP III	KPN Research		nl					tel	•					
Russian speech database	STC/St. Petersburg	ELRA	ru	54	35	-	-	mic			•		[16]	
Gandalf	KTH		sv	48	38	51	32	tel	•		•	•	[45,49]	•
VeriVox	KTH, VeriVox-proj.		sv	-	-	50	0	mic	•		•		[35]	

Table 2. An overview on 36 databases sorted according to language. 'Cat. 1' include speakers recorded in more than one session, and 'Cat. 2' speakers a single session. '96 survey' indicates details for a database are included in the 1996 database survey. 1) recording is under way – the number of speakers will increase; 2) No. of category 1-speakers includes targets in the SPIDRE subset. More speakers are recorded in multiple sessions; 3) Each speaker was recorded through 10 different telephone handsets without passing the signal through a telephone circuit; 4) SUBTV: 'short utterance-based talker verification'; 5) Broertjes contains Polyphone-like sessions so the two can be used to complement each other.

8. REFERENCES

[1] Altincay H., Demirekler M. (1998). "Combination of model confusion information on the basis of Dempster-Shafer theory for speaker identification", COST250, Aarhus meeting, Oct. 8-9.

[2] Altincay H., Demirekler M. (1999). "On the use of supra model information from multiple classifiers for robust speaker identification", Eurospeech'99, Budapest, Hungary, September 5-9, pp. 971-974.

[3] Ambikairajah E., Hassel P. (1996). "Speaker identification on the Polycost database using line spectral pairs", Proc. of COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 47-54.

[4] Ariyaeeinia A.M., Sivakumaran P. (1995). "Speaker verification based on the orthogonalisation technique" Proc. the IEE European Convention on Security and Detection (ECOS '95), Brighton, UK, No. 445, May, pp. 101-105.

[5] Ariyaeeinia A. M., Sivakumaran P. (1997). "Analysis and comparison of score normalisation methods for text-dependent speaker verification", Eurospeech'97, Rhodes, Greece, Sept. 22-25, pp. 1379-1382.

[6] Besacier L., Luettin J., Maître G., Meurville E. (1999). "Experimental evaluation of text-independent speaker verification on laboratory and field test databases in the M2VTS project", Eurospeech'99, Budapest, Hungary, September 5-9, pp. 751-754.

[7] Bimbot F., Hutter H.P., Jaboulet C., Koolwaaij J., Lindberg J., Pierrot J.B. (1998). "An overview of the CAVE project research activities in speaker verification", Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C), Avignon, France, April 20-23, pp 215-220.

[8] Bimbot F., Blomberg M., Boves L., Chollet G., Jaboulet C., Jacob B., Kharroubi J., Koolwaaij J., Lindberg J., Mariethoz J., Mokbel C., Mokbel H. (1999). "An overview of the PICASSO project research activities in speaker verification for telephone applications", Eurospeech'99, Budapest, Hungary, September 5-9, pp. 1963-1966.'

[9] Boves L., Bogaart T., Bos L. (1994). "Design and recording of large data bases for use in speaker verification and identificaiton", Proc. ESCA workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland, April 5-7, pp. 43-46.

[10] Campbell Jr. J., Reynolds D. (1999). "Corpora for the evaluation of speaker recognition systems", ICASSP'99, Phoenix, USA, March 15-19, pp. 829-832.

[11] Cole R., Noel M., Noel V. (1998). "The CSLU speaker recognition corpus", ICSLP'98, Sydney, Australia, November 30– December 4, pp. 3167-3170.

[12] Demirekler M. (1999). "In information theoretic approach for weight estimation in combining multiple probabilistic classifiers", COST250, Porto meeting, March 4-5.

[13] Durou G., Magrin-Chagnolleau I. (1997). "Speaker identification on the Polycost database using the fusion of histogram classifiers method", COST250, Martlesham meeting, April 10-11.

[14] Durou G. (1998). "Cross-language speaker identification on the Polycost database", COST250, Aarhus meeting, Oct. 8-9.

[15] Durou G. (1999). "Time-frequency principal components of speech applied to speaker recognition", COST250, Athlone meeting, June 28-29.

[16] ELRA, European Language Resources Association, http://www.icp.grenet.fr/ELRA.

[17] Falcone M. (1995a). "The speaker recognition database collected at FUB: a report on gained experience", COST250, Rome meeting, Jan. 16-17.

[18] Falcone M. (1995b). "Acoustic characterisation of SIVA the Muser database: some preliminary remarks", COST250, Rome meeting, Jan. 16-17.

[19] Falcone M., Contino U. (1995). "Acoustic characterisation of speech databases: an example for speaker verification", Proc. of The XIIIth Intl. Congress of Phonetic Sciences (ICPhS), Stockholm, Sweden, August, pp. 290-293.

[20] Falcone M., Gallo (1996). "The SIVA speech database for speaker verification: description and evaluation", ICSLP'96, Philadelphia, USA, October, pp. 1902-1905.

[21] Falcone M., Melin H., Ariyaeeinia A. (1999). "Speaker recognition assessment and dissemination: activities in COST250 working group 4", Proceedings of COST250 workshop on Speaker Recognition in Telephony, Rome, Italy, November 10-12.

[22] Furui S. (1986). "Research on individuality features in speech waves and automatic speaker recognition techniques", Speech Communication, Vol 5, pp 183-197.

[23] García-Plaza L., Fernández C. (1996). "Multi-model verification for teleservices and security applications", Proc. COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 7-8.

[24] Gauvain J.L., Lamel L.F., Prouts B. (1995). "Experiments with speaker verification over the telephone", Eurospeech'95, Madrid, Spain, pp. 651-654.

[25] Genoud D., Chollet G. (1999). "Deliberate imposture: a challenge for automatic speaker verification systems", Eurospeech'99, Budapest, Hungary, September 5-9, pp. 1971-1974.

[26] Godfrey J., Holliman E., McDaniel J. (1992). "Switchboard: telephone speech corpus for research and development", ICASSP'92, San Fransisco, USA, March 23-26, pp. 517-520.

[27] Godfrey J., Graff D., Martin A. (1994). "Public databases for speaker recognition and verification", Proc. ESCA workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland, April 5-7, pp. 39-42.

[28] Hassell P., Ambikairajah E. (1997). "Text-independent speaker identification using the Polycost database", COST250, Martlesham meeting, April 10-11.

[29] Hennebert J. (1997). "Polycost annotation status report", COST250, Martlesham meeting, April 10-11.

[30] Hennebert J., Melin H., Petrovska D., Genoud D. (1998). "Polycost 1.0: status of the work", COST250, Ankara meeting, April 1-2.

[31] Hennebert J., Melin H., Petrovska D., Genoud D. (1999). "Polycost: a telephone-speech database for speaker recognition", *to appear in:* Speech Communication.

[32] Hernando J. (1998). “Verification experiments for text-independent speaker identification using frequency filtered spectral energies”, COST250, Aarhus meeting, Oct. 8-9.

[33] Hernando J., Nadeu C. (1998). “Speaker verification on the Polycost database using frequency filtered spectral energies”, ICSLP’98, Sydney, Australia, November 30 – December 4, pp. 129-132.

[34] Höge H., Tropf H.S., Winski R., van den Heuvel H., Haeb-Umbach R. & Choukri K. (1997). “European Speech databases for telephone applications”, ICASSP’97, Munich, Germany, pp. 1771-1774.

[35] Karlsson I., Banziger T., Dankovicova J., Johnstone T., Lindberg J., Melin H., Nolan F., Scherer K. (1998). "Speaker Verification with Elicited Speaking-styles in the VeriVox project", Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C), Avignon, France, April 20-23, pp. 207-210. *Extended version to appear in:* Speech Communication.

[36] Kuitert M. (1996). “Introducing SESP: a realistic database for speaker verification”, Proc. COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 27 (abstract).

[37] Lamel L.F., Gauvain J.L. (1998). “Speaker verification over the telephone”, Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C), Avignon, France, April 20-23, pp. 76-79.

[38] LDC, Linguistic Data Consortium, USA, http://www.ldc.upenn.edu.

[39] Lindberg B. (1996). “Specification of speaker verification databases in SpeechDat (II)”, COST250, Stockholm meeting, June 18-19.

[40] Lindberg J., Melin H., Lundin F., Sundberg E. (1996). “Survey of databases”, COST250, Stockholm meeting, June 18-19. *Also in:* COST250 Final Report, 1999.

[41] Lindberg J., Blomberg M. (1999). "Vulnerability in speaker verification, a study of possible technical impostor techniques", Eurospeech’99, Budapest, Hungary, September 5-9, pp. 1211-1214.

[42] Magrin-Chagnolleau I., Durou G. (1999). “Time-frequency principal components of speech: application to speaker identification”, Eurospeech’99, Budapest, Hungary, September 5-9, pp. 759-762.

[43] Mariéthoz J., Mokbel C. (1999). “Synchronous alignment”, Research Report #99-06, IDIAP, Martigny, Switzerland.

[44] Melin H. (1995). “Design and recording of the Gandalf database”, COST250, Lausanne meeting, Dec. 11-12.

[45] Melin H. (1996). “Gandalf – a Swedish telephone speaker verification database”, ICSLP’96, Philadelphia, USA, October, pp. 1954-1957.

[46] Melin H., Lindberg J. (1996). “Guidelines for experiments on the Polycost database” (Version 1.0), Proc. COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 59-69.

[47] Melin H. (1998). “Optimizing variance flooring in HMM-based speaker verification”, COST250, Ankara meeting, April 1-2.

[48] Melin H., Koolwaaij J.W., Lindberg J., Bimbot F. (1998). “A comparative evaluation of variance flooring techniques in HMM-based speaker verification”, ICSLP’98, Sydney, Australia, November 30 – December 4, pp. 1903-1906.

[49] Melin H., Lindberg J. (1999). “Variance flooring, scaling and tying for text-dependent speaker verification”, Eurospeech’99, Budapest, Hungary, September 5-9, pp. 1975-1979.

[50] Melin H. (1999). “Databases for speaker recognition: working group 2 final report”, *In:* COST250 Final Report.

[51] Melin H., Lindberg J. (1999). “Guidelines for experiments on the Polycost database ” (Version 2.0), *In:* COST250 Final Report.

[52] Mella O. (1994). “Extraction of formants of oral vowels and critical analysis for speaker characterization”, Proc. ESCA workshop on Automatic Speaker Recognition, Identification and Verification, Martigny, Switzerland, April 5-7, pp. 193-196.

[53] Messer K., Matas J., Kittler J., Luettin J., Maitre G. (1999). “XM2VTSDB: the extended M2VTS database”, Proc. of Audio and Video-based Biometric Person Authentication (AVBPA), Washington DC, USA, March 22-23.

[54] Nataf A. (1996). “Definition of environmental and speaker specific coverage for SDB”, SpeechDat, Technical Report LE2-4001-SD1.2.3, October. http://www.speechdat.com.

[55] Nordström T., Melin H. (1998). “Polycost experiment results”, COST250, Ankara meeting, April 1-2.

[56] Nordström T., Melin H., Lindberg J. (1998). “A comparative study of speaker verification systems using the Polycost database”, ICSLP’98, Sydney, Australia, November 30 – December 4, pp. 1359-1362. Also presented in COST250, Aarhus meeting, Oct. 8-9, 1998.

[57] OGI, Oregon Graduate Institute, Center for Spoken Language Understanding, USA, http://www.cse.ogi.edu/cslu.

[58] Olsen J., Lindberg B. (1999). “Algorithms & parameters for speaker recognition: activities in COST250 working group 3”, Proceedings of COST250 workshop on Speaker Recognition in Telephony, Rome, Italy, November 10-12.

[59] Ortega-García J., González-Rodríguez J., Marrero-Aguiar V., Díaz-Gómez J.J., García-Jiménez R., Lucena-Molina J., Sánchez-Molera J.A.G. (1998). “AHUMADA: a large speech corpus in Spanish for speaker identification and verification”, ICASSP’98, Seattle, USA; pp. 773-776.

[60] Ortega-García J. (1999). “Speaker recognition oriented ’AHUMADA’ large speech corpus”, COST250, Porto meeting, March 4-5.

[61] Pawlewski M., Downey S. (1996). “Channel effects in speaker recognition”, Proc. of Institute of Acoustics (I.O.A.) Vol. 18, Part 9, pp. 115–122. *Also in:* Proc. COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 39-46.

[62] Petrovska D., Hennebert J., Genoud D., Chollet G. (1996). “Semi-automatic HMM-based annotation of the Polycost database”, Proc. COST workshop on Applications of Speaker Recognition Techniques in Telephony, Vigo, Spain, November, pp. 23-26.

[63] Petrovska D., Hennebert J., Melin H., Genoud D. (1998). "Polycost: a telephone-speech database for speaker recognition", Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C), Avignon, France, April 20-23, pp. 211-214.

[64] Pigeon S., Vandendorpe L. (1997). "The M2VTS multimodal face database", *In:* Proc. Audio and Video-based Biometric Person Authentication (AVBPA), Springer LNCS, Bigün et al., Eds, 1997.

[65] Reynolds D. A. (1997). "HTIMIT and LLHDB: speech corpora for the study of handset transducer effects", ICASSP-97, Munich, Germany, pp. 1535-1538.

[66] Rodríguez-Liñares L. (1999). "Estudio y Mejora de Sistemas de Reconocimiento de Locutores mediante el Uso de Información Verbal y Acústica en un Nuevo Marco Experimental", Ph.D. Thesis, ETSE de Telecomunicación, University of Vigo, Spain.

[67] Rodríguez-Liñares L., García-Mateo C., Alba-Castro J.L. (1999). "On the use of neural networks to combine utterance and speaker verification systems in a text-dependent speaker verification task", Eurospeech'99, Budapest, Hungary, pp. 1003-1006.

[68] Satué-Villar A., Faúndez-Zanuy M (1999). "On the relevance of language in speaker recognition", Eurospeech'99, Budapest, Hungary, September 5-9, pp. 1231-1234.