Curriculum vitæ

Name :	Dominique Alain Genoud
Birth date :	June 26, 1962.
Origin :	Bagnes, VS, Switzerland.
Languages :	French (mother tongue), English, Spanish, German.
Family status :	married, one child.
Private address :	CP 1508, CH-1951 Sion VS, Switzerland
voice	+ 41 - 27 - 288 52 24
Professional address :	IDIAP
	CP 592, CH-1920 Martigny, VS, Switzerland
voice	+ 41 - 27 - 721 77 26
e-mail	`Dominique.Genoud@idiap.ch`

Education

1999 Ph.D. thesis at the Swiss Federal Institute of Technology, section systems of communication in January 1999, Lausanne, Switzerland.
Title: *Recognition and transformation of speakers.*

1991-1992 Postgraduate course at the Computer Science Department, EPFL: Biological and Artificial Neural Networks.

1989 Eng. Dipl. of the Swiss Federal Institute of Technology Lausanne. (Microtechnic).

Experiences

1999 Hired by Nuance Communications, Menlo Park, California, as Rearcher in the R&D group.

1999 Postdoc researcher at International Computer Science Institute, Berkeley CA, working on speaker recognition, speaker detection and speaker adaptation for large vocabulary speech recognizers.

1995-1999 Researcher at IDIAP, Martigny, Switzerland, in speaker recognition domain, involved in diverse European projects (Cost, Telematics, ACTS), assumed a leading role on the Swiss side of the projects.

1994-1995 Research engineer at IDIAP, responsible for research & development of speaker recognition systems for the Swiss telecom and for diverse other telecom related companies.

1993-1994 Round the world trip.

1992-1993 Team manager of the research & development department of CIMTEC (Swiss competence center for quality insurance, Sion, Switzerland), also acting as teacher of methods for product development at Ecole d'ingénieur du Valais (EIV), Sion, Switzerland.

1991-1992 Project leader at EIV, Sion, Switzerland, responsibility of industrial projects in high precision measurement systems, (mechanical and optical systems).

1989-1991 Member of the research and development team in NAGRA-Kudelski, Cheseaux, Switzerland, involved in the elaboration of a new data format including software and hardware development for professional digital audio tape recorders.

Research interests

- Speech and audio processing
- Bio-engineering processing
- Machine learning, non-linear modeling
- Digital signal processing

Research activities

I started my research activites at IDIAP Switzerland as reserach engineer with the responsability of developping a text dependent speaker recognition system for the Swiss telecom. The building and the developpment of this system is the starting point of my PhD thesis.

Purpose of the thesis

My PhD thesis tries to understand how to analyse, decompose, model and transform the vocal identity of a human when seen through an automatic speaker recognition application. The errors of the basic operating speaker recognition application were analysed, and from the deficiencies and mistakes noticed in the running application, some observations were be made which implied a re-evaluation of the characteristic parameters of a speaker, and to reconsider some parts of the automatic speaker recognition chain.

speaker ID by his voice

In order to determine what are the characterising parameters of a speaker, these are extracted from the speech signal with an analysis and synthesis harmonic plus noise model (H+N). The analysis and re-synthesis of the harmonic and noise parts indicate those which are speech or speaker dependent. It is then shown that the speaker discriminating information can be found in the residual of the subtraction from the original signal of the H+N modeled signal. Then, a study of the impostors phenomenon, essential in the tuning of a speaker recognition system, is carried out. The impostors are simulated in two ways: first by a transformation of the speech of a source speaker (the impostor) to the speech of a target speaker (the client) using the parameters extracted from the H+N model. This way of transforming the parameters is efficient as the false acceptance rate grows from 4% to 23%. Second, an automatic imposture by speech segment concatenation is carried out. In this case the false acceptance rate grows to 30%. A way to become less sensitive to the spectral modification impostures is to remove the harmonic part or even the noise part modeled by the H+N from the original signal. Using such a subtraction decreases the false acceptance rate to 8% even if transformed impostors are used.

Modeling a speaker with few data

To overcome the lack of training data – one of the main cause of modeling errors in speaker recognition – a decomposition of the recognition task into a set of binary classifiers is proposed. A classifier matrix is built and each of its elements has to classify word by word the data coming from the client and another speaker (named here an anti-speaker, randomly chosen from an external database). With such an approach it is possible to weight the results according to the vocabulary or the neighbours of the client in the parameter (acoustic) space. The output of the matrix classifiers are then weighted and mixed in order to produce a single output score. The weights are estimated on validation data, and if the weighting is done properly, the binary pair speaker recognition system gives better results than a state of the art HMM based system.

Statistical corrections

In order to set a point of operation (i.e. a point on the COR curve) for the speaker recognition application, an *a priori* threshold has to be determined. Theoretically the threshold should be speaker independent when stochastic models are used. However, practical experiments show that this is not the case, as due to modeling mismatch the threshold becomes speaker and utterance length dependant. A theoretical framework showing how to adjust the threshold using the local likelihood ratio is then developed.

Decision Fusion

As a concluding chapter of my PhD thesis a last modeling error correction method using decision fusion is proposed. Some practical experiments show the advantages and drawbacks of the fusion approach in speaker recognition applications.

Simultaneous speech and speaker decoding

Since my arrival at ICSI I used as basis the hybrid large vocabulary speech recognition system available at ICSI, and I modified the MLP which provide the frame by frame a posteriori probability estimation in order to simultaneousely perform a speaker verification (using a log-likelihood-ratio-like score for each of the registered speaker), and a speech recognition (giving the phoneme probability).

The first exploratory results, using the DARPA Broadcast News-97 database, show that the performances of the speech recognition remains the same and the speaker recognition performances are around 6% of Equal Error Rate, which is very encouraging regarding the difficulty of the task.

I Also worked in the speaker adaptation of the large vocabulary speech recognizers, this adaptation shows a real improvement of the speech recognition performance.

Technical skills & knowledge

Computer programming

- assemblers (68xxx, DSP 56k), conventional langages (Pascal, Fortran, C), OO-languages (Pascal, C++, java), other (perl, SQL, industrial languages, ...).

Operating systems

- UNIX (linux,SunOS,HPux), Win32/NT.

Electronic

- Digital electronic design, digital signal processing, FPGA programming, simulation languages.

Other

- knowledge in optic and image processing.

Non technical skills

Good knowledge of technical projects leading, like teamwork, like to manage people, have a strategic view.

Publications

International journals

- **An overview of the CAVE project research activities in speaker verification**, *submitted to SpeechCom journal*, 1998.
(with F. Bimbot, M.B lomberg, L. Boves, H.P. Hutter, C. Jaboulet, J. Koolwaaij, J. Lindberg and J.B. Pierrot)

- **Polycost : a telephone-speech database for speaker recognition**, *submitted to SpeechCom journal*, 1998.
(with D.Petrovska, J.Henneberg, and H. Melin)

- **Acoustic-Labial Speaker Verification**, *in Pattern Recognition Letters*, 18:9, pp 853-858, 1997.
(with P. Jourlin, J. Luettin and H. Wassner)

International conferences

- **Client/world model synchronous alignement for speaker verification**, *to appear in proc of EUROSPEECH-99*, Budapest, 1999.
(with J. Mariethoz, F. Bimbot and C. Mokbel)
- **Deliberate imposture: a challenge for automatic speaker verification systems.**, *to appear in proc of EUROSPEECH-99*, Budapest, 1999.
(with G. Chollet)
- **Speech pre-processing against intentional imposture in speaker recognition**, *in proc. of ICSLP-98*, Sidney, 1998.
(with G. Chollet)
- **Voice-B project**, *in proc. of IVTTA-98*, Torino, 1998.
(with G. Caloz, C. Jaboulet, J. Mariéthoz, A. Glaeser)
- **Voice transformation, a tool for imposture of speaker verification**, *in proc. of International Phonetic Science conference IPS98*, Washington, 1998.
(with G. Chollet)
- **A comparison of a priori threshold setting procedures for speaker verification in the CAVE project**, *in proc ICASSP 98*,Seatle, 1998.
(with J.B. Pierrot, J. Lindberg, J. Koolwaaij, H.P. Hutter, M. Blomberg and F. Bimbot)
- **Text dependent speaker verification using binary classifiers**, *in proc ICASSP 98*, Seatle, 1998.
(with M. Moreira and E. Mayoraz)
- **POLYCOST : a telephone-speech database for speaker recognition**, *in proc RLA2C* Avignon, 1998.
(with D. Petrovska, J. Hennebert and H. Melin)
- **Integrating Acoustic and Labial Information for Speaker Identification and Verification**, *in EUROSPEECH97*, Rhodos, 1997.
(with P. Jourlin, J. Luettin and H. Wassner)
- **Likelihood ratio adjustment for the compensation of model mismatch in speaker verification**, *in EUROSPEECH97*, Rhodos, 1997.
(with F. Bimbot)
- **Acoustic-Labial Speaker Verification**, *in Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Crans, 1997. (with P. Jourlin, J. Luettin and H. Wassner)

- **Secured vocal access to telephone servers**, *in Proceedings of IVTTA 1996 IEEE Third Workshop Interactive Voice Technology for Telecommunications Applications*, Baskin Ridge NJ-USA, 1996.
  (with O. Bornet, G. Chollet, J.L. Cochard and A. Constantinescu)
- **Polycost Database**,
  in Joint meeting cost249-250, Stockholm, 1996.
  (with J. Hennebert and H. Melin)
- **Combining methods to improve speaker verification decision**,
  ICSLP 96, Philadelphia, 1996.
  (with F. Bimbot, G. Gravier, and G. Chollet)
- **Semi-automatic HMM-based annotation of the PolyCOST Database**,
  in Application of speaker recognition techniques in telephony-COST250 workshop, Vigo-Es, 1996.
  (with D. Petrovska-Delacretaz, J. Hennebert and G. Chollet)

French-speaking Conferences

- **Proposition d'une stratégie de fusion de données à trois niveaux pour la vérification d'identité**,
  in proc. of Journées d'Etudes sur la Parole, Martigny, 1998. (with P. Verlinde, G. Gravier and G. Chollet)
- **Système de vérification du locuteur dépendante du texte utilisant des classificateurs binaires**,
  in proc. of Journées d'Etudes sur la Parole, Martigny, 1998.
  (with Miguel Moreira and Eddy Mayoraz)
- **Amélioration des performances de vérification du locuteur par combinaison de méthodes**,
  Journées d'études sur la parole JEP96, Avignon, 1996.
  (with F. Bimbot, G. Gravier, and G. Chollet)