Extended Curriculum vitæ

Name: **LANCHANTIN**
First Name: **Pierre**

Date of birth: **23th june 1976 in Paris**

Nationality: **French**

Marital status: **single**

E-mail: **pierre.lanchantin@gmail.com**

Web page: **http://recherche.ircam.fr/equipes/analyse-synthese/lanchant**

This Curriculum Vitæ includes the following sections :

Education

09/2002-12/2006 **PhD in Statistical Signal Processing** - *CITI department - Institut National des Télécommunications (INT)*

Title : Triplet Markov chains and Unsupervised signals segmentation / Chaînes de Markov Triplets et Segmentation Non Supervisée de Signaux.

PhD thesis from Institut National des Télécommunications.

Keywords : Bayesian estimation, Hidden Markov Models (chains and trees), Pairwise and triplet Markov models, Bayesian network, Multivariate and Multiresolution image segmentation, Expectation Maximization, Iterative Conditional Estimation, Dempster-Shafer theory of evidence, fuzzy set theory, non stationary processes, Gaussian processes, long memory processes.

Mention : très honorable.

Defended on the 5th december 2006 in front of the following committee :

Christophe COLLET	Reviewer	ENSPS/ULP, Strasbourg
Patrick PEREZ	Reviewer	IRISA, Rennes
Henri MAITRE	Examiner	ENST, Paris
Igor NIKIFOROV	Examiner	UTT, Troyes
Wojciech PIECZYNSKI	Director	INT, Evry
Fabien SALZENSTEIN	Co-supervisor	ULP, Strasbourg

09/2001-09/2002 **Master's Degree in System Optimisation and Safety (DEA OSS)** - *INT, UTT, URCA*

Major: Signal processing and decision theory.

With honors

Keywords: Logistic, Risk management, System diagnostic, Decision theory in Signal and Image

09/2000-09/2001 **Master's Degree in Acoustics, Signal Processing and Computer science applied to Music (DEA ATIAM)** - *University of Paris VI, IRCAM, ENST*

With honors

Keywords: Acoustics(general and musical), Audio-numerical signal processing, Computer Science for Music.

09/2000-09/2002 **Master's Degree in Telecommunications (french Grande Ecole Telecom INT)** - *INT*

With honors

Keywords: Signal processing, Computer Science, Probability and statistics, graph optimization, numerical analysis, Information theory, numerical communications, optical communications, Network-TCP/IP.

Major in statistical image processing during the last year.

09/1997-06/2000 **Bachelor and first year of Master of Science in Physics** - *University of Evry-val-d'Essonne*

With honors

Keywords: Statistical Physics, Optics, Electromagnetism, Relativity, Quantum Mechanics, Electronics, Numerical Analysis

09/1995-06/1997 **Audio engineer diploma** - *School of Audio Engineering, Paris, France*

A two years degree in sound audio technics

Keywords: studio and live recording technics, sonorisation, Acoustics, mastering.

09/1989-06/1995 **Secondary Education** - *Versailles Academy*

General education in Sciences

Professional experience

01/2010-06/2011 **Researcher in FEDER Angelstudio project** - Analysis-Synthesis team - *IRCAM, Paris, France*

- Research on **Voice Conversion** (Dynamic Model Selection) and development of a system based on GMM modeling of the joint distribution of source and target acoustic features.
- Implementation of a **one-to-many voice conversion** system based on a canonical eigenvoice model estimated by SAT for fast Adaptation.
- The aim of VC in this project was to convert the voice of a commercial text-to-speech system (TTS) to the voice of the user using few sentences.

- Research and development of a **HMM-based speech synthesis** system for French based on HTK/HTS toolbox with high level syntactical features with N. Obin.

07/2008-01/2010 **Researcher in ANR Affective avatars project** - Analysis-Synthesis team - *IRCAM, Paris, France*

- Research on **Voice conversion** (reduction of the conditional variance) and development of a Voice conversion system based on GMM modeling of the joint density of source and target acoustic features.
- Research and development of a **HMM-based speech synthesis** system for French based on HTS including a new excitation model (SVLN) proposed by G. Degottex.

01/2007-06/2008 **Researcher in ANR VIVOS project** - Analysis-Synthesis team - *IRCAM, Paris, France*

- Research and development of a **speech segmentation** system based on HTK and on the french phonetizer LIAPHON to automatically extract the language structure at different levels (phone, word, phrase, paragraph) and to align it on the speech audio signal.
- Multiple pronunciation are taking into account during the alignment using a constrained phonetic graph build from the text.
- A **confidence measure** is computed for manual correction.
- The aligned linguistic structure was used in the project by ircamCorpusTools, a corpus manager tool similar to festival, which was developed for Unit selection TTS.

09/2005-09/2006 **Teaching assistant (ATER) for Master's degree students:** - *Paris XI University, Orsay, France*

- Teaching: C language, UNIX, numerical analysis, multimedia (coding), system and network, final projects supervision.

01/2003-09/2005 **Teaching assistant for Master's degree students:** - *INT, Evry, France*

- Teaching: Introduction to Statistics, Algorithmics and C language, statistical methods in Image processing, final projects supervision.

09/2002-12/2002 **Invited researcher** - Ocean Systems Laboratory - *Heriot-Watt University, Edinburgh, UK*

- Study and evaluation of SONAR images segmentation algorithms.

03/2002-09/2002 **Master Training course in statistical RADAR image segmentation** - /RDTA - TBU Radar Development - DTO - *THALES Air Defence, Bagneux, France*

- Study of statistical radar image segmentation algorithm in the application field of Doppler cartography in order to reduce false alarm in RADAR detection

03/2001-09/2001 **Master Training course in non-linear Mechanics** - UER de Mécanique - *ENSTA, Palaiseau*

- Analytical and numerical study of the temporal response of a circular plate involving a set of internal resonances in the context of non linear vibration.

01/2000-03/2000 **Training course in non-linear Optics** - Photonics and nanostructures laboratory - *CNET, Bagneux*

- Simulation of the propagation of a gaussian laser beam in a non-linear medium (C++)

Responsabilities and collective charges

Supervision — Supervision of 2 students final projects (Master telecom INT) on the study of statistical models for unsupervised segmentation of signals. Supervision of 2 groups of students in Master on a Voice over IP project (RTP/RCTP, SIP protocols).

Projects — Participation in the preparation of ANR (french national research agency) projects (VOCO, Affective Avatar) and FEDER (Respoken, Angelstudio).

Reviewing — Reviewer for Speech communication and IEEE Transactions on Audio, Speech and Language Processing.

Computer skills

OS — Unix, Linux, Windows, Mac OS X.

Languages — C, C++, UML, JAVA, HTML, LaTeX/ Script: Bash, awk, Perl, Lisp

Engineering	Matlab, PV-WAVE
Libraries	SPTK, HTK, HTS
Music	ProTools, LogicAudio, Max/MSP, Ableton live, Melodyne, Reason.
Office	Microsoft Office, Star Office,

Languages

French	Mother tongue.
English	Fluent. First Certificate in English (Grade B).
Spanish	Basic knowledge.

Interests/activities

Music	Pianist, bass player in jazz and R&B bands. Electronic Music creations.
Sport	Boxing, kali (Filipino martial art), running, swimming.

Referees

Dr Axel Roebel

Capacity in which known: Head of the analysis Synthesis Team
address: IRCAM, analysis synthesis team, 1, place Igor Stravinsky, 75004 Paris, France
email: axel.roebel@ircam.fr
phone number: + (33) 1 44 78 48 45

Pr Xavier Rodet

Capacity in which known: my direct supervisor at Ircam
address: IRCAM, analysis synthesis team, 1, place Igor Stravinsky, 75004 Paris, France
email: rod@ircam.fr
phone number: + (33) 1 44 78 48 68

Pr Wojciech Pieczynski

Capacity in which known: my PhD thesis director
address: CITI Laboratory, Telecom SudParis, 9 Rue Charles Fourier, 91011 Evry Cedex, France
email: wojciech.pieczynski@it-sudparis.eu
Phone number: + (33) 1 60 76 44 25

Research

Introduction

My research area is *statistical signal processing*. My main topics of research include **statistical modeling of signals**, **speech processing** and their **applications to Music**. My research has focused initially on the generalization of statistical models for signals, especially *Hidden Markov Models* (HMM). I studied during my PhD (T) models called Triplet Markov models that generalize the classical HMM with applications in image segmentation. I then directed my research toward speech processing, working on the segmentation of speech signals into phones, voice conversion, language models and HMM-based speech synthesis. My research studies on speech are interdisciplinary as they combine statistical modeling of signals, natural language processing and their application to music.

First, I present my PhD studies on unsupervised segmentation of signals based on triplet Markov models. Then, I present my work on speech processing, performed in the analysis and synthesis team at IRCAM.

Pairwise and Partially Pairwise Markov chains

The principle of a **pairwise Markov chain** is to assume that the joint distribution of observed and hidden processes is that of a Markov chain. In the case of a **partially pairwise Markov chain**, one assumes directly the Markovianity of the law of the hidden process conditionally to the observations. In this context, one of my contributions was to propose the **expectation-maximization** (EM) algorithm in the case of pairwise Markov chains. I also studied with W. Pieczynski a special case of partially pairwise Markov chain applied to the segmentation of **Gaussian processes with long correlation** [C3]. Experiments on synthetic data gave significant improvements compared to conventional models where the noise is a long correlated one while giving similar performance when the noise was independent. Nevertheless, the proposed method of parameters estimation was only valid for the centered case, which prevented us from testing the model on real images. So we continued our work with J. Lapuyade to refine our methods and make possible the unsupervised segmentation of Gaussian processes which are not necessarily centered [A2].

Triplet Markov chains

Pairwise Markov models can be extended to **triplet Markov models** [C1]. The principle is to add one, or even several, auxiliary process(es) as the joint distribution of the triplet "hidden process, auxiliary processes, observed process" is that of a Markov chain. These very general models allow to palliate another limitation of conventional models which is to assume that the joint distribution is stationary. Indeed, by introducing an auxiliary process controlling changes in transition matrices of the process, we have shown the effectiveness of such a model in situations where the joint distribution of the hidden process and observation is **nonstationary** [C2] and we proposed algorithms for estimating parameters of a the considered Markov chain. This model was applied to the segmentation of synthetic and real images. A first observation is that this model does allow the consideration of different regimes, resulting in improved quality of segmentation in the case of images with both extensive homogeneous areas and areas with fine details. A second observation is that it is also possible to obtain a realization of the auxiliary process by the MPM estimator. This type of representation can be very useful, especially in segmentation of textures that can be precisely modeled by auxiliary processes.

Data fusion

We have also studied with W. Pieczynski, as part of the study of triplet Markov models, the possibilities for extending the classical probabilistic model to an "evidential" model, with the posterior probability of the hidden process given by the **Dempster-Shafer fusion** [A1, AF1, CF2] . We then applied this evidential model to the segmentation of **nonstationary processes**. The main interest of our approach was to show that although the Dempster-Shafer fusion destroys Markovianity in the context of the hidden evidential chain, Bayesian segmentation is still possible via the triplet Markov chain approach.

A last of my contributions during my PhD thesis was to extend the fuzzy Markov chains previously studied by F. Salzenstein, to fuzzy Markov trees. The **fuzzy segmentation** was initially proposed to take account of imprecision on a site belonging to a thematic area. Thus, in a fuzzy signal cohabit homogeneous areas ("hard" clusters) with fuzzy areas representing intermediate sites which may belong to several hard clusters. The originality of these models is characterized by the fact that their distribution has both a discrete and a continuous components, the discrete component being formed by Dirac masses representing the weight assigned to each cluster lasts and the continuous component corresponding to the fuzzy classes (Lebesgue measure). We have proposed a **multisensor fuzzy hidden Markov tree** that we applied to the segmentation of astronomical images [CF3].

HMM-based Speech Segmentation

During the ANR Vivos, I proposed and developed in collaboration with A. C. Morris and X. Rodet, the software *ircamAlign* [C4]. This system of **segmentation of speech signals into phones** is based on the HTK library. It is based on a particular modeling of speech signals by hidden Markov chains used in speech recognition. This modeling can be viewed as a special case of Markov chain triplet $T = (U, X, Y)$ where U is the language model, X is the process of evolution of spectral features in time (sub-states of the HMM of each phoneme) and Y is the process of the observations (cepstral coefficients).

Based on this observation, if the textual transcription exists, the distribution of the process U can be defined as that of a Markov chain whose topology is a graph constructed from the phonetic text giving the different pronunciations and possible connections. Many options are available for creating this graph. It is thus possible to allow the omission or repetition of words, the insertion of short pauses or paraverbal sounds like breathing or lip noises for which specific models have been learned. When the text is not available, such as in the case of a spontaneous speech signal, the distribution of U is defined as being that of a bigram or tri-gram learned on a selected French text set.

A set of multispeaker French models have been learned from the corpus BREF80. On the other hand, a confidence index based on posterior probabilities is calculated for each phone to facilitate a possible manual correction of segmentation results.

During the segmentation, the structure of speech (syllables, words, breath groups) is extracted from transcription and aligned to the speech signal in order to build databases of units for text-to -Speech (TTS) (C7) by units selection. *ircamAlign* is used by ircamTTS and ircamCorpusTools [AF2] which is a management system database of speech units. On the other hand, *ircamAlign* is used in the ANR Rhapsody project for developing reference corpus of spontaneous speech in French. Finally, ircamAlign has been used by composers at IRCAM. Note that a real-time version has subsequently been developed by J. Bloit and implemented in MaxMSP.

HMM-based Speech Synthesis

The principle of **HMM-based speech synthesis** developed by the Nagoya Institute of Technology (Nitech) is the joint modeling of the spectrum (vocal tract), the fundamental frequency (source) and durations for each phoneme in context by a hidden Markov chain. During the synthesis, a macro-model is built from the concatenation of the HMMs corresponding to the phones in the context of the phonetic sequence to synthesize. The durations of the states are initially generated and then the trajectories of spectral parameters are estimated from a specific algorithm for spectral parameters generation taking into account the dependency between static and dynamic parameters. One advantage of this method compared to the synthesis of speech by units selection is that it only requires the storage of model parameters. It also allows precise control of the characteristics of the synthesis. The disadvantages of this type of synthesis are artifacts in the synthesized voice due to the glottal source modeling and the lack of natural due to the low variability of the prosody. To overcome these shortcomings we used the **separation of vocal tract and glottal source separation method** proposed by G. Degottex in [C5]. On the other hand, we have shown with N. Obin the improvement made by using **high-level syntactical features** [C6] and the possibilities for the synthesis of **speaking style** for different types of **discourse genres** in [C11].

Voice Conversion

The principle of **Voice conversion** is to transform the signal from the voice of a source speaker, so it seems to have been issued by a target speaker. Conversion techniques I studied at IRCAM are based on **Gaussian Mixture Models** (GMM). Typically, the joint distribution of acoustic source and target characteristics, modeled by a GMM, is estimated from a parallel corpus consisting of synchronous recordings of source and target speakers. The conversion function is then given by the conditional expectation to the acoustic characteristics of the source. My studies have focused both on the definition of the transformation function on its application to improve the quality of converted speech. Thus, all-pole modeling of the spectral envelope has been improved by the True-Envelope technics that enhances the quality of the synthesis and the characterization of the residual from the speaker. On the other hand, the use of the covariance matrix of the conditional distribution to the acoustic characteristics of the source allows a renormalization of the transformed characteristics in order to improve the quality of the converted signal. Finally, during the AngelStudio project, I proposed a method for **Dynamic Model Selection** (DMS [C8, C10]) which consits in using several models of different complexity and to select the most appropriate model for each frame of analysis during the conversion. The results of voice conversion obtained are very encouraging. Thus, it appears that the "personality" of the target speaker is well reproduced after processing and that the source speaker has largely disappeared. The main difficulty that remains is some degradation of sound quality of voice. However, other ways of improvements we are currently investigating [C13] are expected to arrive at a usable quality, real-time, even for very demanding applications, such as artistic applications.

Publications

Refereed journal publications

[A3] **P. Lanchantin**, J. Lapuyade-Lahorgue and W. Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise, *Signal Processing*, No. 2, Vol. 91, pp 163-175, Feb 2011.

[A2] **P. Lanchantin**, J. Lapuyade-Lahorgue and W. Pieczynski, Unsupervised segmentation of Triplet Markov chains hidden with long memory noise, *Signal Processing*, No. 88, Vol. 5, pp 1134-1151, May 2008.

[A1] **P. Lanchantin** and W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chains using evidential priors, *IEEE Transactions on Signal Processing*, Vol. 53, No. 8, pp 3091-3098, 2005.

Refereed journal publications in French

[AF2] G. Beller, C. Veaux, G. Degottex, N. Obin, **P. Lanchantin** et X. Rodet, IrcamCorpusTools : Plateforme Pour Les Corpus de Parole, *Traitement Automatique des Langues*, Vol. 49, No. 3, 2008

[AF1] **P. Lanchantin** et W. Pieczynski, Chaînes et arbres de Markov évidentiels avec applications à la segmentation des processus non stationnaires, *Traitement du Signal*, Vol. 22, No. 2, 2005.

Conference Proceedings

[C10] **P. Lanchantin** and X. Rodet, Objective Evaluation of the Dynamic Model Selection for Spectral Voice Conversion, *ICASSP2011*, accepted, Prague, Czech Republic, 2011

[C9] **P. Lanchantin**, S. Farner, C. Veaux, G. Degottex, A. Roebel, X. Rodet, A short review on voice transformations at IRCAM, *Proc. of the First International Workshop on Performative Speech and Singing Synthesis*, Vancouver, Canada, March 2011

[C8] **P. Lanchantin** and X. Rodet, Dynamic Model Selection for Spectral Voice Conversion, *Interspeech'10*, Makuhari, Japan, 2010

[C7] C. Veaux, **P. Lanchantin** and X. Rodet, Joint Prosodic and Segmental Unit Selection for Expressive Speech Synthesis, *7th Speech Synthesis Workshop (SSW7)*, Kyoto, Japan, 2010

[C6] N. Obin, **P. Lanchantin**, M. Avanzi, A. Lacheret-Dujour and X. Rodet, Toward Improved HMM-Based Speech Synthesis Using High-Level Syntactical Features, *Speech Prosody*, Chicago, USA, 2010

[C5] **P. Lanchantin**, G. Degottex and X. Rodet, A HMM-Based Synthesis System Using a New Glottal Source and Vocal-Tract Separation Method, *ICASSP'10*, Dallas, USA 2010,

[C4] **P. Lanchantin**, A. C. Morris, X. Rodet, C. Veaux, Automatic Phoneme Segmentation with Relaxed Textual Constraints, in E. L. R. A. (ELRA) (ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 2008.

[C3] W. Pieczynski, **P. Lanchantin**, Restoring hidden non stationary process using triplet partially Markov chain with long memory noise, *IEEE Workshop on Statistical Signal Processing (SSP 05)*, July 17-20, Bordeaux, France, 2005.

[C2] **P. Lanchantin** and W. Pieczynski, Unsupervised non stationary image segmentation using triplet Markov chains, *Advanced Concepts for Intelligent Vision Systems (ACIVS 04)*, Aug. 31-Sept. 3, Brussels, Belgium, 2004.

[C1] W. Pieczynski, D. Benboudjema and **P. Lanchantin**, Statistical image segmentation using Triplet Markov Fields, *SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.

Submitted

[C13] **P. Lanchantin**, N. Obin and X. Rodet, Extended Conditional GMM and Covariance Matrix Correction for Real-Time Spectral Voice Conversion, *submitted to Interspeech 2011,* Florence, Italy, 2011

[C12] N. Obin, **P. Lanchantin**, A. Lacheret-Dujour and X. Rodet, Reformulating Prosodic Break Model into Segmental HMMs and Information Fusion, *submitted to Interspeech 2011,* Florence, Italy, 2011

[C11] N. Obin, **P. Lanchantin**, A. Lacheret-Dujour and X. Rodet, Discrete/Continuous Modelling of Speaking Style in HMM-based Speech Synthesis: Design and Evaluation, *submitted to Interspeech 2011,* Florence, Italy, 2011

Conference Proceedings in French

[CF4] X. Rodet, G. Beller, N. Bogaards, G. Degottex, S. Farner, **P. Lanchantin**, N. Obin, A. Roebel, C. Veaux, F. Villavicencio, Transformation et Synthèse de la voix parlée et de la voix chantée, *Actes du Colloque de Rentrée du Collège de France*, Octobre 2008.

[CF3] **P. Lanchantin**, F. Salzenstein, Segmentation d'Images Multispectrales par Arbre de Markov caché Flou, *Actes du Colloque GRETSI'05*, 6-9 septembre, Louvain-la-Neuve, Belgique, 2005.

[CF2] **P. Lanchantin** et W. Pieczynski, Arbres de Markov Triplet et théorie de l'évidence, *Actes du Colloque GRETSI'03*, 8-11 septembre, Paris, France, 2003.

[CF1] C. Touzé, **P. Lanchantin**, A. Chaigne et O. Thomas, Transferts d'énergie par couplage modal : étude d'un cas particulier, *Actes du sixième congrès français d'Acoustique*, Lille, Avril 2002.

PhD Thesis

[T] **P. Lanchantin**, *Chaînes de Markov Triplets et Segmentation Non Supervisée de Signaux*, PhD thesis from Institut National des Télécommunications, december 2006

Press

- Chercheur à l'Ircam : sur la bonne voix, M. Davène, *Keyboards Recording*, août 2009.

Talks

- **P. Lanchantin**, Conversion d'Identité Vocale, *Journée prospective IRCAM*, dec. 2010.
- C. Veaux, **P. Lanchantin**, N. Obin, J. Bloit, Exploitation d'enregistrements et de corpus de Parole, *Journée prospective IRCAM*, dec. 2008.
- G. Beller, J. Bloit, G. Degottex, S. Farner, **P. Lanchantin**, N. Obin, A. Roebel, C. Veaux, F. Villavicencio, X. Rodet, Travaux actuels de l'Ircam autour de la voix, *ILPGA*, Paris, 4 avril 2008.
- **P. Lanchantin**, A. C. Morris, X. Rodet, ircamAlign : système d'étiquetage et d'alignement de signaux de parole, *IRCAM research and technology seminars*, sep. 2007.
- J. Bloit, **P. Lanchantin**, X. Rodet, Décodage dans un flux : reconnaissance et alignement de parole, *IRCAM research and technology seminars*, sep. 2007.
- **P. Lanchantin**, Modèles Triplets et Théorie de l'Evidence, *Séminaire LSIIT* (ENSPS, Strasbourg), mai 2006.
- D. Benboudjema, N. Brunel, **P. Lanchantin**, W. Pieczynski, Traitement statistique des images : principes de modélisation pour la segmentation, *Journée portes ouvertes Recherche (JPOR 05) de l'INT* : présentation des activités de recherche en traitements statistiques d'images du laboratoire CITI, 23-24 mars 2005.
- **P. Lanchantin**, Segmentation non-supervisée de processus cachés non stationnaires par modèle de chaîne de Markov triplet, *Séminaire CITI*, Février 2005.

Internal report

- **P. Lanchantin**, ircamAlign : Système d'étiquetage et d'alignement de signaux de parole, *rapport Interne Ircam*, Université Pierre et Marie Curie, 2007