Daniel Moraru

Né le 22.03.1977 à Bucarest, Roumanie (28 ans)
Nationalité : Roumaine

14B Bd. De Metz, Ap 21
35000 Rennes
Tel : 06 88 71 40 89
Mail : dmoraru@irisa.fr
daniel.moraru@gmail.com

CANDIDATURE POST-DOC
Doctorant 3ème année Traitement du Signal

Résume des compétences

Techniques	Traitement du Signal, Traitement de la Parole, Electronique	
	C/C++ , Visual Basic, Perl, HTML, Assembleur INTEL, ARM, Sockets Réseau	
	Matlab, Windows, Linux	
Linguistiques	Français (couramment)	Roumaine (langue maternelle)
	Anglais (couramment)	

Formation

20 Dec. 2004	Docteur en traitement du signal, spécialité Signal Image Parole Télécoms (SIPT), Thèse soutenue à l'Institut National Polytechnique de Grenoble
2000 – 2001	DEA Signal Image Parole Télécoms (SIPT), Institut National Polytechnique de Grenoble
1995 – 2000	Diplôme d'ingénieur Université "Politehnica", Faculté d'Electronique et Télécommunications, Bucarest, Roumanie

Position antérieure

A partir de 02.2005	Chercheur post-doctoral à l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA/INRIA), Rennes, France
Pendant 2004	Partie de l'équipe qui a développé ALIZE, une librairie logiciel libre pour la vérification et identification du locuteur (http://www.lia.univ-avignon.fr/heberges/ALIZE/)
2001 - 2004	Thèse en traitement de la parole, "Segmentation en locuteurs de documents audio et audiovisuels : application à la recherche d'information multimédia" au Laboratoire CLIPS (Communication Langagière et Interaction Personne-Système), Grenoble Allocataire de Recherche INP Grenoble, Ecole Doctorale EEATS

Activités de recherche

Voir annexe détaillée :

- Traitement de la parole

- Modélisation statistique pour la reconnaissance du locuteur
- Segmentation parole / non-parole pour des documents audio
- Segmentation en locuteur et reconnaissance du locuteur pour des journaux télévisés
- Adaptation non-supervisé au locuteur pour les systèmes de reconnaissance de la parole continue à grande vocabulaire
- Participations très honorables aux évaluations organisées par NIST (National Institute of Standards and Technologies) sur la segmentation audio (SpRec 02, RT 03, RT 04) et sur la segmentation vidéo (TREC 11, 12, 13)

Publications

Voir annexe détaillée:

- 1 revue internationale en cours de soumission
- 9 conférences internationales avec actes et comite de lecture
- 3 conférences nationales avec actes et comite de lecture

Activités d'enseignement

2003	24h	TD	Traitement du Signal	UFR-IMA	(Responsable Anne Guérin-Dugué)
2003	18h	TD	Architecture Logicielles et Matérielles	ISTG PolyTech	(Responsable Pascal Sicard)
2003	18h	TD	Architecture Logicielles et Matérielles	UFR-IMA	(Responsable Anne Guérin-Dugué)
2002	24h	TP	Informatique scientifique	ISTG Polytech	(Responsable Patricia Ladret)
2002	48h	TP	Traitement du Signal	ENSERG	(Responsable Alice Caplier)
2002	-	-	Jury d'entretien Eiffel	ISTG Polytech	(Responsable Henry Ettinger)
2001	24h	TP	Traitement du Signal	ENSERG	(Responsable Alice Caplier)
2001	44h	TD	Bureautique	UPMF	(Responsable Nathalie Denos)

Stages

2002 Stage "Estimation de nombre de locuteurs dans un document audio",Laboratoire DDL ("Dynamique du Langage"), Institut des Sciences de l'Homme Lyon

2000 Stage "Serveur de synthèse vocale pour le projet européen NESPOLE!", Laboratoire CLIPS – IMAG, Grenoble

1999 Stage "Utilisation de la méthode de cohortes pour la reconnaissance du locuteur", Graphco International, Bucarest, Roumanie

Annexe 1

Activités d'enseignement

Electronique, Traitement de Signal

2003	Traitement du Signal	Ufr-IMA 1er année Filière MIAGE	Introduction au traitement du signal en temps continu, Série de Fourier, Transformée de Fourier, Convolution, Filtrage, Modulation en Amplitude	24 h	TD
2002	Traitement du Signal	ENSERG 2ème année Filière Télécom	Transformée de Fourier Discrète, Analyseur du Spectre, Récepteur Radio	48 h	TP
2001	Traitement du Signal	ENSERG 2ème année Filière Télécom	Transformée de Fourier, Modulation de Fréquence	24 h	TP

Informatique

2003	Architectures Logicielles et Matérielles	ISTG Polytech 1er année Filière RICM	Architecture microprocesseurs, Langage machine, Micro Langage	18 h	TD
2003	Architectures Logicielles et Matérielles	Ufr-IMA 2ème année Filière MIAGE	Architecture microprocesseurs, Langage machine, Micro Langage	18 h	TD
2002	Informatique scientifique	ISTG Polytech 1er année Licence EEA	Langage C, Initiation à la programmation	24 h	TP

Autres

2002	Jury d'entretien Eiffel	ISTG Polytech étudiants BAC+2	Entretien pour admission en première année	26	entretiens
2001	Bureautique	UPMF 1er année DGES	Windows, Word, Excel	44 h	TD

Total

- 104 heures TD
- 96 heures TP

Annexe 2
Activités de recherche

Mots clés:

- Indexation automatique de documents multimédia
- Traitement de la parole
- Modèles statistiques
- Reconnaissance du locuteur
- Recherche d'information sur données multimédia

Stage DEA:

Titre: Indexation de signaux en locuteurs

En indexation automatique de documents multimédia, il existe différentes tâches liées au traitement de signaux audio et de parole : sous-titrage automatique d'un film, classement des éléments de la bande son d'un film, séparation parole / musique, détection de changement de locuteurs. Le sujet de mon stage était de réaliser un système d'indexation en locuteurs c'est à dire reconnaître qui parle, et à quel moment, sur un document sonore contenant la voix de plusieurs personnes.

Dans un premier temps, j'ai approfondi les aspects théoriques concernant mon sujet. Les signaux qui devront être indexés contiennent seulement deux locuteurs et sont divisés en deux catégories : signaux simulés (obtenus en concaténant des phrases issues d'une base de données de parole) et signaux réels (des vrais dialogues via IP, toujours deux locuteurs par signal). Aucun modèle acoustique statistique de locuteurs ou de parole n'est disponible au départ. L'indexation est faite en deux étapes : la segmentation (détection des changements de locuteurs) et le regroupement (les segments issus de la première étape sont réunis en deux classes). Outre le système d'indexation proprement dit, ce travail a apporté une contribution concernant les procédures d'évaluation dans ce nouveau domaine qu'est l'indexation audio.

Thèse:

Titre : Segmentation en locuteurs de documents audio et audiovisuels : application à la recherche d'information multimédia

Début : Octobre 2001

Soutenue : le 20 Décembre 2004

Ma thèse a commencé en octobre 2001 dans le Laboratoire CLIPS au sein du Groupe d'Etude sur l'Oral et le Dialogue (GEOD) sous l'encadrement de monsieur Laurent Besacier.

La thèse porte sur l'indexation de documents audio par locuteurs, le but étant d'améliorer les systèmes de recherche d'information dans des très grandes bases de données multimédia mais aussi d'améliorer les performances de systèmes de reconnaissance de la parole continue tel que "RAPHAEL", le système de reconnaissance du laboratoire.

Mon travail a commencé en première année avec la mise en place d'un système d'indexation issu de l'état de l'art, système basé sur l'adaptation par locuteur de modèles statistiques génériques indépendants du locuteur.

Au cours de ma première année de thèse j'ai eu notamment la possibilité de collaborer avec le laboratoire LIA (Laboratoire d'Informatique d'Avignon) qui travaille sur le même thème de recherche. Ensemble nous avons participé avec nos systèmes d'indexation aux évaluations internationales SpRec 2002 (Speaker Recognition) organisées par NIST (National Institut of Standards and Technologies). A cette occasion le système du laboratoire a été déclaré meilleur système d'indexation sur la tâche "Meeting" (enregistrements de réunions).

En deuxième année de thèse j'ai commencé une partie plus exploratoire. Au cours des mois de janvier et février j'ai effectué un séjour à Lyon au laboratoire DDL (Dynamique du Langage). A cette occasion j'ai travaillé à l'amélioration du système d'indexation existant par une méthode d'estimation du nombre de locuteurs présents dans un document audio.

Je précise ci-dessous pour les personnes qui veulent en savoir plus sur cette méthode, quelques détails techniques. La méthode est basée sur le critère d'information Bayesien pénalisé. Le principe consiste à calculer la vraisemblance du document analysé étant donné les modèles statistiques utilisés $\log L(X/M)$, et à pénaliser cette vraisemblance par la complexité des modèles. Il s'agit de déterminer le nombre de locuteurs NSp qui maximise :

$$BIC(M)=\log L(X/M)-\lambda\frac{m}{2}N_{Sp}\log N_X$$

Dans cette formule, M est le modèle qui réunit tous les modèles de locuteurs présents dans le document, X est la totalité des trames acoustiques du document audio, Nx est la taille du document audio, m est un paramètre qui exprime la complexité du modèle et λ est un paramètre de réglage.

Cette méthode donne des très bons résultats, notamment sur des enregistrements de grande taille. En effet, cette étape est indispensable dans le cas de certains documents comme les journaux télévisés ou le nombre de locuteurs est inconnu. La méthode a été appliquée avec succès aux évaluations NIST 2003 et 2004.

La partie expérimentale de mon travail de recherche a été réalisée sur des enregistrements réels de type journaux télévisés ou radio: CNN, NBC, France Inter, Radio

France International de durée variable: de 10 minutes à 1 heure et aussi avec un nombre de locuteurs variable de 13 à 40, d'où l'intérêt de l'estimation du nombre de locuteurs.

Pendant ce deux années j'ai aussi participé en collaboration avec une autre équipe du laboratoire qui travaille sur la recherche d'information, aux évaluations TREC (Text Retrieval Conference) organisées aussi par NIST. Les évaluations TREC sont dédiées à la recherche d'information et donnent la possibilité de fusionner des informations sur un même document provenant de sources différentes (audio, vidéo, texte). Les évaluations TREC m'ont permis d'évaluer l'apport d'un système d'indexation audio dans le cadre de la recherche d'informations et de me lancer vers une des parties exploratoires de ma thèse.

La participation aux évaluations NIST a donné lieu à plusieurs publications dans des conférences internationales et aussi ma permis de collaborer avec d'autres chercheurs dans le domaine et dans des domaines connexes (traitement de la vidéo, recherche d'information).

Partie Exploratoire :

Dans le cadre des l'évaluations NIST, l'indexation audio est faite en aveugle : sans aucune information *a priori* sur le documents analysés autre que le signal audio lui même. Cependant entre différents types de documents analysés (journaux télévisés, réunions, enregistrements téléphoniques) il existe des informations *a priori* exploitables. Par exemple : dans le cadre des enregistrements téléphoniques les interventions de locuteurs sont plus courtes que dans le cas des enregistrements de journaux télévisés; dans le cas des enregistrements de journaux télévisées le présentateur est généralement connu d'avance, etc. J'ai donc exploré l'apport que les informations *a priori* sur les documents audio pourront apporter à mon système. Ce travail a donné lieu à une publication dans une conférence internationale très reconnue dans le domaine.

Dans la dernière partie de la thèse, nous nous sommes intéressé à l'application de notre système à une tâche d'indexation d'une grande collection de documents vidéo. Nos travaux de recherche ont été faits dans le cadre des évaluations TREC (Text REtrieval Conference) organisées également par NIST.

Notre insertion dans ce domaine nouveau s'est faite via la participation officielle du laboratoire à certaines tâches proposés dans les évaluations TRECVID 2002 et 2003. De cette expérience, nous avons défini nos propres tâches de recherche, intéressantes d'un point de vue *locuteur*, dans le but de répondre à deux questions principales : est-ce que la segmentation en locuteurs peut apporter des informations utiles pour l'indexation de documents vidéo ? Et, inversement, est-ce que les informations extraites à partir du canal vidéo peuvent être utiles pour améliorer le résultat d'un système de segmentation en locuteurs ?

Concernant l'apport de l'information vidéo sur une tâche purement audio de segmentation en locuteurs, les résultats obtenus ont montré une réduction d'environ 3 % en absolu de l'erreur grâce à un système de segmentation en locuteurs "audio-visuel". Ce système utilise la détection de frontières de plans vidéo pour améliorer l'étape de détection de changements de locuteurs de notre système de segmentation en locuteurs. Un plan vidéo

est défini comme une séquence continue de trames vidéo similaires d'un point de vue visuel. Les évaluations TREC utilisent le terme "*continuous camera shots*", un plan (shot) étant le résultat des trois actions suivantes : démarrer la caméra vidéo, filmer une séquence vidéo, arrêter la caméra vidéo. Les moments du démarrage et l'arrêt de la caméra vidéo constituent les frontières d'un plan vidéo. Dans le cas de transmissions télévisées un changement de caméra constitue aussi une frontière.

A partir d'un résultat de segmentation en locuteurs, nous avons proposé une méthode originale pour la recherche de plans vidéos contenant le présentateur d'un journal télévisé. L'originalité de l'approche est due à l'absence d'un modèle probabiliste du locuteur cible, présent dans tout système de suivi de locuteur classique. Le présentateur est identifié uniquement à partir du résultat du système de segmentation en locuteurs et avec quelques règles empiriques issues de connaissances sur la structure d'un journal télévisé.

Les dernières expérimentations présentent une approche multimodale pour la segmentation en histoires de documents vidéo. Dans cette approche, nous avons montré que les informations *locuteur* obtenues avec les méthodes présentées dans cette thèse, permettent d'aider à améliorer un système de découpage plus sémantique d'un document vidéo.

Annexe 3

Publications importantes

Revues internationales en cours de soumission :

- "Approaches in broadcast news speaker segmentation", S.Meignier, D. Moraru, C. Fredouille, J.-F. Bonastre, L. Besacier, Computer Speech and Language, acceptée et en phase d'apparition

Conférences internationales :

- "Experiments on Speaker Tracking and Segmentation on Radio Broadcast News", D. Moraru, M. Ben, G. Gravier, EuroSpeech 2005, Lisbon, Portugal, September 2005
- "Speaker and Story Segmentation Using Audio-Video Information", D. Moraru, L. Besacier, G. Quenot, S. Ayache, Speech Technology and Human-Computer Dialogue 2005, Cluj-Napoca, Romania, May 2005
- "Using a priori information for speaker diarization", D.Moraru, L.Besacier, E.Castelli A Speaker Odyssey 2004, Toledo, Spain, Mai 2004
- "ELISA NIST RT 03 Broadcast News Speaker Diarization Experiments", D. Moraru, S. Meignier, C. Fredouille, L. Besacier, J-F Bonastre, A Speaker Odyssey 2004, Toledo, Spain, Mai 2004
- "The ELISA Consortium approaches in broadcast news speaker segmentation during the nist 2003 rich transcription evaluation", D. Moraru, S. Meignier, C. Fredouille, L. Besacier, J-F Bonastre, ICASSP 2004, Montreal, Canada, Mai 2004
- "Benefits of prior acoustic segmetation for automatic speaker segmentation", S. Meignier, D. Moraru, C. Fredouille, L. Besacier, J-F Bonastre, ICASSP 2004, Montreal, Canada, Mai 2004
- "The ELISA consortium approaches in speaker segmentation during the NIST 2002 speaker recognition evaluation", D. Moraru, S. Meignier, L. Besacier, J-F Bonastre, I. Magrin-Chagnolleau. ICASSP 2003, Honk-Kong, China, Mai 2003
- "CLIPS-IMAG at TREC-11 : Experiments in Video Retrieval", Georges M. Quénot, Daniel Moraru, Laurent Besacier, and Philippe Mulhem , 11th Text Retrieval Conference, Gaithersburg, MD, USA, Novembre 2002
- "Speech Translation for French in the NESPOLE! European Project", L. Besacier, H. Blanchon, Y. Fouquet, J.P. Guilbaud, S. Helme, S. Mazenot, D. Moraru, D. Vaufreydaz, Eurospeech 2001, Aalborg, Danemark, Septembre 2001

Conférences nationales

- "Premiers pas du CLIPS sur les données d'évaluation ESTER", R. Lamy, D. Moraru, B. Bigi, L. Besacier, JEP 2004, Fès, Maroc, Avril 2004
- "Segmentation selon le locuteur: les activités du Consortium ELISA dans le cadre de Nist RT03", D. Moraru, S. Meignier, C. Fredouille, L. Besacier, J-F Bonastre, JEP 2004, Fès, Maroc, Avril 2004
- "Separation en locuteurs de conversations via IP", D. Moraru, L. Besacier, JEP 2002, Nancy, France, Juin 2002