

Les systèmes du LIA pour les tâches de segmentation et de suivi: SES, SRL, SVL

N. Scheffer, D. Istrate, C. Fredouille, J.-F. Bonastre,

LIA
UAPV, Avignon, FRANCE

(nicolas.scheffer,dan.istrate,corinne.fredouille,jfb)@lia.univ-avignon.fr

Abstract

Ce papier décrit les systèmes de segmentation et de suivi du Laboratoire d'Informatique d'Avignon (LIA) présentés lors de la campagne d'évaluation ESTER 2005. Cette campagne a permis une première validation de ces systèmes étant données les performances obtenues tout à fait satisfaisantes. Les systèmes présentés sont entièrement basés sur la plateforme open-source ALIZE, dédiée à la Reconnaissance Automatique du Locuteur.

1. Introduction

La première campagne ESTER d'Evaluation des Systèmes de Transcription Enrichie d'émissions Radiophoniques a eu lieu en Janvier 2005, proposant d'une part des tâches de transcription automatique (TTS et TTR) et d'autres part des tâches de segmentation et suivi d'événements telles que le suivi d'événements sonores (SES), la segmentation et le regroupement en locuteurs (SRL) et le suivi de locuteurs (SVL). Par ailleurs, une tâche de détection des entités nommées a été également proposée. Le Laboratoire Informatique d'Avignon (LIA) a participé activement à cette campagne, proposant un système pour chacune des tâches proposées.

L'objectif de ce papier est de présenter un descriptif des systèmes du LIA pour les tâches de segmentation et de suivi ainsi que les performances de ces derniers sur les corpus de développement - ESTER DEV Phase II noté *Dev* dans la suite du papier - et d'évaluation - ESTER TEST Phase II noté *Eva* - disponibles durant la campagne.

Tous ces systèmes ont été entièrement développés sur la plateforme ALIZE dédiée à la Reconnaissance Automatique du Locuteur [1]. Ce toolkit a été développé par le LIA dans le cadre du projet Technolangue/AGILE/ALIZE, financé par le ministère français de la recherche. Des informations sur la plateforme ALIZE sont disponibles à l'adresse suivante : <http://www.lia.univ-avignon.fr/heberges/ALIZE>. Les codes sources et programmes utilisés au cours de cette évaluation seront prochainement disponibles à cette même adresse, en licence de type GPL.

Les sections 2, 3 et 4 sont dédiées aux descriptions des différents systèmes liés aux tâches SES, SRL et SVL respectivement. Leurs performances sont présentées et discutées dans les sections 5, 6 et 7. Une conclusion générale sur la participation à cette évaluation est finalement fournie en section 8.

2. Description du système SES

2.1. Tâche SES

La tâche de suivi d'événements sonores consiste à détecter de manière automatique les portions de signal dans lesquels un événement particulier est présent. Les événements à suivre sont ici la musique (en présence ou non de parole) et la parole (en présence ou non de musique).

Le système de suivi d'événements sonores (SES) du LIA repose sur une segmentation en macro-classes acoustiques, utilisée par d'autres systèmes du LIA tels que : le système de suivi de locuteurs (tâche SVL), le système de segmentation/regroupement en locuteurs (tâche SRL) et le système de transcription automatique (tâches TRS et TTR).

2.2. Système de segmentation en macro-classes

Système de base

Le système de segmentation en macro-classes acoustiques repose sur une segmentation hiérarchique à deux niveaux :

- Durant la première étape, une segmentation parole+musique, parole bande large, parole bande étroite (téléphone) et non parole est appliquée sur le signal (une émission complète). Quatre modèles *MS*, *S*, *T* et *M* sont utilisés lors de cette segmentation, représentant respectivement les conditions parole+musique, parole bande large, parole bande étroite et non parole. Ce processus de segmentation repose, ici, sur un décodage Viterbi, appliqué sur un HMM ergodique à quatre états (modèles *MS*, *S*, *T* et *M*).
- La deuxième étape se focalise sur une détection du genre, durant laquelle chaque segment parole (*MS*, *S*, *T*) issu de la phase précédente est étiqueté soit homme soit femme. Cette détection s'appuie sur le même principe de décodage que l'étape précédente, associé à des modèles dépendants du genre et de la classe acoustique du segment traité. Ainsi, les modèles *SF* et *SH* représentent respectivement des segments de parole (parole bande large et parole+musique) produits par des femmes et des hommes, *TF* et *TH* des segments de parole téléphonique femmes et hommes. Finalement, la sortie du décodage est soumise à un ensemble d'heuristiques (règles sur la durée minimale des segments par exemple) permettant d'affiner la segmentation.

Spécifications du systèmes

Tous les modèles utilisés lors des différentes phases de segmentation sont des mixtures de gaussiennes (GMM)

caractérisées par des matrices de covariance diagonales. Tous les modèles sont caractérisés par 64 gaussiennes. Ce nombre de composantes a été fixé empiriquement. Chaque modèle GMM est estimé sur un sous-ensemble de données issus des 80 heures d'apprentissage de la campagne ESTER - Phase 2.

Lors de la paramétrisation, le signal de parole est caractérisé par un vecteur de 39 coefficients, composé de 12 MFCC (estimés toutes les 10ms sur une fenêtre d'analyse de type Hamming de 20ms), de la log-énergie ainsi que des dérivées premières et secondes. Aucune normalisation des paramètres n'est appliquée. Aucun traitement particulier n'est utilisé pour la parole téléphonique.

2.3. Système de suivi d'événements sonores

Le système de suivi en événements sonores repose uniquement sur un ré-étiquetage des sorties de la première étape du système de segmentation en macro-classes acoustiques (la deuxième étape - détection du genre - étant uniquement nécessaire pour d'autres systèmes du LIA (tâches TRS, TTR, SRL et SVL). Les segments étiquetés *MS* et *M* prennent l'étiquette "music", les segments étiquetés *S*, *T* et *MS* deviennent "speech".

L'objectif de la classification en macro-classes était principalement de détecter la musique seule en tant que segments non parole afin de l'écarter des traitements réalisés en aval de la chaîne (transcription, segmentation en locuteur, ...) et non de suivre cet événement (en présence ou non de parole). Par conséquent, aucune attention particulière n'a été apportée en vue d'améliorer les résultats du suivi de l'événement "music" en présence de parole.

3. Description du système SRL

3.1. Tâche SRL

La tâche de segmentation et regroupement en locuteur (SRL) consiste à découper un document audio faisant intervenir plusieurs locuteurs en des segments homogènes ne contenant qu'un seul locuteur (phase de segmentation) et à regrouper ces derniers dès lors qu'ils appartiennent au même locuteur (phase de regroupement ou clustering). Généralement, cette tâche est accomplie sans aucune information a priori sur l'identité et le nombre de personnes intervenant dans le document sonore.

Le système SRL du LIA repose essentiellement sur l'utilisation d'un modèle de Markov caché évolutif, appelé E-HMM, représentatif de la conversation entre les locuteurs [2, 3, 4, 5]. Chaque état de cet E-HMM caractérise un locuteur particulier du document sonore et ses transitions modélisent les changements entre locuteurs. L'E-HMM étant ergodique, tous les changements de locuteurs sont autorisés.

Le processus SRL est ici basé sur une approche itérative durant laquelle les locuteurs sont détectés et ajoutés un à un au E-HMM à chaque nouvelle itération. Dans ce contexte, toutes les informations disponibles sont utilisées à chaque étape et remises en cause à l'étape suivante. L'originalité principale de cette approche tient au fait que le processus de segmentation et de regroupement sont réalisés conjointement.

3.2. Principe de base : E-HMM

Durant le processus de segmentation et regroupement en locuteurs, un modèle de Markov caché (HMM) est généré sur la base d'une approche itérative qui détecte et ajoute à chaque

itération un nouvel état, ie un nouveau locuteur au modèle de conversation évolutif (E-HMM). Ce processus itératif de segmentation et regroupement en locuteurs est réalisé au travers de 4 étapes :

Etape 1-Phase d'initialisation

Un premier modèle de locuteur, nommé L_0 , est estimé sur l'ensemble des données sonores à traiter¹. A cette étape, la segmentation est modélisée par un HMM à un seul état correspondant au locuteur L_0 , représentant l'ensemble des locuteurs présents dans les données sonores (et non encore détectés comme tels).

Etape 2-Ajout d'un nouveau locuteur

Un nouveau locuteur est détecté parmi les segments étiquetés L_0 , grâce à une technique de sélection de données (décrite dans la section suivante). Ces dernières seront utilisées pour estimer un nouveau modèle de locuteur. Un nouvel état, nommé L_x , est ainsi rajouté au modèle HMM, suivi d'une mise à jour des probabilités de transition entre états (voir la section suivante pour le détail de cette mise à jour). Finalement, les données sélectionnées lors de la détection du locuteur L_x sont étiquetées comme telles dans la segmentation courante (et perdent par conséquent leur ancienne étiquette L_0).

Etape 3-Adaptation/Décodage

L'objectif de cette étape est de détecter tous les segments appartenant au nouveau locuteur L_x en remettant en cause la segmentation courante et par conséquent l'appartenance des segments aux autres locuteurs (en particulier L_0) déjà présents dans le modèle HMM. Dans ce sens, tous les modèles de locuteurs sont tout d'abord réestimés par adaptation suivant la segmentation courante. Une phase de décodage par l'algorithme Viterbi est ensuite lancée afin de produire une nouvelle segmentation. Ce processus Adaptation/Décodage est réitéré jusqu'à stabilisation de la segmentation entre deux phases successives.

Etape 4- Validation du modèle de locuteur et critère d'arrêt

A l'issue de l'étape précédente, la segmentation courante est examinée pour déterminer si l'ajout du nouveau locuteur est pertinent dans le processus de SRL. Deux heuristiques sont appliquées dans cette décision :

- le nouveau locuteur doit avoir retenu dans la nouvelle segmentation un ensemble de segments dont la durée doit dépasser celle du segment initialement sélectionné ou une durée minimale d_{seg} ; cette règle a pour objectif d'éliminer très tôt les modèles de locuteurs mal estimés².
- le nouveau locuteur doit avoir retenu dans la nouvelle segmentation un ensemble de segments dont la durée doit dépasser 10% du locuteur L_0 ; cette règle doit permettre de sélectionner (autant que possible) les locuteurs dominants de L_0 .

Si le locuteur ne répond pas à ces deux règles, il est supprimé du modèle HMM.

Le critère d'arrêt est atteint dès lors qu'aucune donnée dans le locuteur L_0 n'est disponible pour le procédé de sélection et l'ajout d'un nouveau locuteur. Dans le cas contraire, le processus reprend à l'étape 2.

¹Différentes classes sonores peuvent être disponibles pour un même document sonore, en particulier si une phase de segmentation en macro classes acoustique est appliquée préalablement

²Cette règle peut-être appliquée plus tôt dans le processus d'Adaptation/Décodage, afin de limiter les temps de calcul.

3.3. Le système complet : système Primaire

Le système complet de segmentation et regroupement en locuteurs du LIA, présenté comme système primaire lors de l'évaluation, se décline en 3 étapes distinctes :

- une segmentation en macro-classes acoustiques (*LIAAcSeg*);
- une phase de segmentation/regroupement (*LIASpkSeg*);
- une phase de resegmentation (*LIASpkReSeg*).

LIAAcSeg

Les documents audio de type Broadcast News peuvent présenter divers environnements acoustiques tels que la musique, la parole sur de la musique, de la parole téléphonique ou non téléphonique, ... Une étude effectuée sur des Broadcast News américaines a montré que la méthode de segmentation et regroupement en locuteurs présentée ici peut bénéficier en termes de performances d'une segmentation préliminaire en macro-classes acoustiques [6].

Cette segmentation en macro-classes acoustiques est identique à celle utilisée pour la tâche SES (voir section 2). Elle fournit pour chaque document sonore en entrée du système SRL quatre classes acoustiques : parole bande large femme *SF*, parole bande large homme *SH*, parole téléphonique femme *TF* et parole téléphonique homme *TH*.

LIASpkSeg

L'approche de segmentation et regroupement en locuteur décrite précédemment est appliquée sur l'ensemble des segments associés à chaque macro-classe acoustique fournie par la segmentation précédente, indépendamment des autres. L'ensemble des segmentations obtenues individuellement sur chaque macro-classe acoustique sont ensuite regroupées pour fournir une segmentation de sortie.

LIASpkReSeg

Cette phase de resegmentation a pour objectif d'affiner la segmentation fournie par l'étape précédente :

- en corrigeant si possible les erreurs de segmentation en macro-classes acoustiques (un locuteur fragmenté à tort en parole téléphone et non téléphone ou suivant le genre);
- en affinant les frontières entre locuteurs ;
- en supprimant les locuteurs considérés comme peu fiables par le système.

Dans ce cadre, seule l'étape 3 de l'approche de segmentation et regroupement en locuteurs du système *LIASpkSeg* est appliquée avec des paramètres différents. En effet, un modèle HMM est construit à partir de la segmentation fournie par l'étape précédente (*LIASpkSeg*). Le processus itératif d'Adaptation/Décodage est alors lancé jusqu'à atteindre une stabilisation de la segmentation courante. A la fin de chaque itération, la segmentation courante est examinée afin de supprimer du HMM les locuteurs (états) dont la durée est inférieure à un seuil d_{reseg} . Il faut remarquer que la resegmentation est lancée sur toutes les macro-classes confondues.

3.4. Système contrastif

Le système contrastif présenté par le LIA lors de la campagne d'évaluation est une variante du système primaire. La seule différence repose sur l'ajout d'une étape supplémentaire lors de la phase de resegmentation permettant d'"affiner" le locuteur L_y (détecté lors de l'étape de seg-

mentation/regroupement) dont le temps de parole est le plus important. En effet, il a été observé que ce locuteur "absorbe" très souvent des segments de parole d'autres locuteurs détectés ou non lors du processus de segmentation/regroupement.

Cette étape supplémentaire consiste à :

1. sélectionner parmi les segments attribués au locuteur L_y un segment en utilisant la même technique de sélection que celle appliquée lors de la phase de segmentation/regroupement ;
2. Adapter (technique d'adaptation de type MAP classique) le modèle associé au locuteur L_y en utilisant uniquement le segment sélectionné ;
3. Appliquer le processus itératif d'Adaptation/Décodage jusqu'à stabilisation de la segmentation de sortie.

Cette étape supplémentaire est ensuite suivie de la phase de resegmentation du système primaire.

3.5. Spécification des systèmes

Paramétrisation

La paramétrisation est réalisée grâce au toolkit SPRO3. Le signal est caractérisé par 20 coefficients cepstraux linéaires (LFCC), issus d'une analyse en bancs de filtres (24 bancs), appliquée toutes les 10ms sur une fenêtre de 20ms. Les vecteurs de 20 LFCC sont augmentés par l'ajout de l'énergie. Aucune normalisation n'est appliquée sur ces vecteurs.

Modélisation des locuteurs/états

Les modèles de locuteurs associés à chaque état du HMM sont des mélanges de gaussiennes (GMM) à 128 composantes (matrices de covariance diagonales) estimées par adaptation d'un modèle générique de parole. Seules les moyennes sont adaptées. Le modèle générique est indépendant du genre et estimé par l'algorithme EM/ML sur une cinquantaine d'heures de parole du set d'apprentissage phase 2 d'ESTER. Deux techniques d'adaptation sont utilisées lors du processus SRL :

- Technique classique d'adaptation de type MAP [7] ;
- Variante de la technique MAP, propre au LIA [3]. L'adaptation des moyennes du modèle de locuteur est obtenue par combinaison pondérée des moyennes du modèle générique (UBM) et de celles estimées à partir des données du locuteur suivant l'équation suivante :

$$\mu_i = \frac{\alpha w_i^{UBM}}{\alpha w_i^{UBM} + (1 - \alpha) w_i^E} \mu_i^{UBM} + \frac{(1 - \alpha) w_i^E}{\alpha w_i^{UBM} + (1 - \alpha) w_i^E} \mu_i^E$$

où w_i^{UBM} et w_i^E sont respectivement les poids pour chaque gaussienne i du modèle générique et du modèle de locuteur, α est une pondération fixée ici à 0.2, μ_i^{UBM} et μ_i^E sont respectivement les moyennes du modèle générique et du modèle de locuteur (estimée par EM/ML).

Lors de l'exécution du système de segmentation/regroupement *LIASpkSeg*, le facteur de régulation nécessaire à l'adaptation classique est fixé à 16 et augmenté à 32 pour l'exécution de la resegmentation *LIASpkReSeg*.

Il est à noter que durant le processus de segmentation/regroupement, la quantité de données disponibles pour l'adaptation des modèles est très variable d'une étape à l'autre et d'un locuteur à l'autre. Par exemple, le locuteur L_0 peut-être fortement avantage par la quantité de données disponibles pour l'apprentissage de son modèle comparé à un nouveau

locuteur "fraîchement" rajouté au HMM (ici juste 6 secondes d'apprentissage lors de la première adaptation). Aussi, une technique de sélection aléatoire de trames a été mise en place pour limiter la quantité de données utilisée lors de l'adaptation, si celle-ci dépasse un seuil fixé empiriquement à 15 minutes.

Modèle HMM de conversation

Le modèle HMM de conversation repose sur l'utilisation de probabilités de transition entre les états. Ces probabilités de transition sont mises à jour à chaque nouvel ajout d'un locuteur au sein du HMM. Cette mise à jour suit les règles suivantes :

- toutes les probabilités de transition d'un état i vers un état j (avec $i \neq j$) sont équiprobables ;
- le rapport γ entre les probabilités intra-état P_{ii} et inter-état P_{ij} (avec $i \neq j$) doit rester constant au fur et à mesure de l'ajout de nouveaux locuteurs. Cette règle se traduit par les équations suivantes :

$$P_{ii} = \frac{\gamma}{\gamma + N - 1}$$

et

$$P_{ij} = \frac{1}{\gamma + N - 1}$$

Le rapport gamma est fixé empiriquement à 9.

Finalement, un facteur multiplicatif f est appliqué sur les probabilités de transition afin de les ramener dans un espace comparable à celui des valeurs de vraisemblance lors de l'application de l'algorithme Viterbi. Le facteur f est ici fixé à 2.

L'estimation des probabilités d'émission diffère de l'approche classique de décodage par Viterbi pour lequel généralement les vraisemblances sont calculées trame à trame. Ici, les vraisemblances sont moyennées sur une fenêtre de 0.3s, résultant en un lissage des changements locaux entre états.

Technique de sélection d'un nouveau locuteur

Comme décrit précédemment, l'approche itérative basée sur l'E-HMM consiste à détecter et rajouter de nouveaux locuteurs au HMM à chaque itération. La détection d'un nouveau locuteur repose sur une technique de sélection d'un segment de parole au sein du locuteur L_0 (ensemble des locuteurs non encore détectés). Sachant que ce segment doit par la suite être attribué au nouveau locuteur pour la phase d'adaptation de son modèle, sa sélection est prédominante dans le processus SRL. La technique de sélection choisie se décline comme suit :

- un segment de durée $d_{selection}$ est sélectionné dans les données associées au locuteur L_0 ; la valeur de $d_{selection}$ doit être suffisamment grande pour estimer, par adaptation, un modèle de locuteur robuste et restreinte pour assurer qu'un seul locuteur intervient dans ce segment ; ce segment est attribué au nouveau locuteur L_x ajouté au HMM, suivi d'une première itération d'Adaptation/Décodage ; La technique d'adaptation utilisée ici est la variante LIA du MAP ; La valeur $d_{selection}$ est fixée ici à 6s.
- A la fin de la phase de décodage, tous les segments attribués au locuteur L_x sont examinés, et seul celui contenant le segment initialement sélectionné est conservé, les autres étant réattribués à L_0 ;
- le processus itératif d'Adaptation/Décodage (Etape 3 du processus SRL) est alors lancé, en utilisant pour chacune des itérations la technique MAP classique (facteur de régulation égal à 16) pour la réestimation de tous les modèles de locuteurs du HMM.

Le changement de technique d'adaptation au cours du processus de sélection permet de prendre en compte la taille variable des données attribuées au locuteur (Variante LIA du MAP dès lors que la quantité de données d'adaptation est faible - données initiales de taille $d_{selection}$ - et MAP classique dès qu'un segment plus grand est susceptible d'être détecté suite à la phase de décodage).

4. Description du système SVL

4.1. Tâche SVL

La tâche SVL consiste à détecter tous les segments de parole correspondant à un locuteur donné, inclus dans une liste de locuteurs à suivre. Contrairement à la tâche SRL, tous les locuteurs à suivre sont connus du système ; des données d'apprentissage sont disponibles pour chacun d'eux permettant l'apprentissage de modèles de locuteurs.

Le système SVL du LIA se décompose en différentes étapes (décrites dans les sections suivantes), basées pour la majorité sur les techniques classiques de détection du locuteur (apprentissage des modèles, tests de détection, normalisation des scores, ...). Seule la phase préalable de segmentation et la décision restent propres à la tâche de suivi de locuteurs.

4.2. Principes de base

4.2.1. Pré-segmentation

Un système de suivi peut reposer sur une pré-segmentation en locuteurs lors de la phase de test. Le système de suivi est ainsi appliqué sur chacun des segments sans se soucier de la détection des frontières entre locuteurs. Ces segmentations peuvent être disponibles (segmentation de références) ou fournies par un système automatique tel que le système SRL (section 3).

De même, l'apprentissage des modèles de locuteurs nécessite, dès lors qu'on traite des sources multi-locuteurs (émissions radiophoniques), une segmentation de référence afin d'identifier les frontières entre locuteurs.

Finalement, les systèmes SVL présentés ici reposent également sur une segmentation en macro-classes acoustiques (fournie par le système SES section 2) fournissant une détection de canal (téléphone/studio) et de genre.

4.2.2. Paramétrisation des signaux

La paramétrisation des signaux se décompose en deux étapes, l'extraction de paramètres et leur normalisation. Il est à noter qu'aucune suppression de trames non-énergétiques (dites de silence) n'a été effectuée.

Paramétrisation MFCC

Le signal est caractérisé par 32 coefficients : 16 coefficients cepstraux sur l'échelle de fréquence de Mel (MFCC) et leurs dérivées respectives.

Normalisation des paramètres

A l'exception du système 5, tous les paramètres ont été normalisés par genre et par condition. Cette normalisation consiste à centrer et réduire la distribution des paramètres de façon à ce que pour chaque condition (HommeStudio, FemmeStudio, ...) et pour chaque fichier, la distribution soit de moyenne 0 et de variance 1.

4.2.3. Apprentissage des modèles du monde

Le type de modèles utilisés pour les modèles sont des modèles à mélange de gaussiennes (GMM). L'initialisation est réa-

lisée en sélectionnant une partie des trames disponibles de façon aléatoire dans le flux pour chacune des composantes. Dans notre cas, nous sélectionnons une trame sur mille.

L'estimation des paramètres des modèles du monde est réalisée grâce à l'algorithme EM. Le principe de sélection aléatoire des trames est réutilisé. Ainsi, une première étape consiste à appliquer 8 itérations d'EM sur 20% des trames. Dix itérations d'EM clôturent le procédé en utilisant cette fois-ci, la totalité des trames disponibles. Une stratégie de contrôle empêche les composantes d'avoir une variance supérieure à 0.5 tout au long de l'apprentissage.

Pour tous les systèmes, nous possédons au moins deux modèles du monde indépendants du genre : un modèle homme et un modèle femme. Des modèles du monde dépendants du canal sont estimés : Téléphone ou Studio dans notre cas. Les modèles utilisés pour tous les systèmes (à l'exception du système contrastif 4) possèdent 1024 composantes. Pour des raisons d'efficacité, un quart des données disponibles pour chaque radio a été utilisé. Il est à noter que les signaux d'apprentissage des locuteurs, ont été utilisés pour estimer les modèles du monde.

Suivant le contexte, l'apprentissage des modèles du monde repose sur une pré-segmentation en macro-classes acoustiques fournie par le système SES.

4.2.4. Modélisation des locuteurs

Les modèles de locuteurs sont adaptés par MAP à partir d'un modèle du monde. Seules les moyennes sont adaptées. Le processus est itératif : quatre itérations utilisant 10% des trames sont effectuées suivies d'1 itération finale utilisant la totalité des trames. Le facteur de régulation adopté a pour valeur 14. Il est à noter que les modèles du monde caractérisant la condition téléphone ont été adaptés de la même manière que les modèles du locuteur ie à partir du modèle du monde caractérisant la condition studio. Le nombre de trames utilisées pour l'apprentissage des modèles a été limité à 120000 (soit 20 minutes). Nous utilisons la segmentation de référence (fichiers mdtm) pour l'apprentissage des modèles de locuteurs.

4.2.5. Test de détection

Un test bayésien sur le rapport de vraisemblance est utilisé comme test de détection du locuteur. De part la quantité importante de tests à effectuer, les vraisemblances sont calculées sur les 10 meilleures gaussiennes sélectionnées grâce au modèle du monde associé. Les tests effectués sont dépendants du genre, ce dernier étant issu de la segmentation en macro-classes acoustiques.

4.2.6. Normalisation des scores

Une normalisation de type T-Norm est utilisée pour normaliser les scores. La moyenne et la variance de la distribution imposteur sont calculées, pour un locuteur donné, sur les scores des autres locuteurs à suivre (en enlevant le score maximum, car nous supposons que le locuteur ayant obtenu un score maximum est un client, et ne doit pas rentrer pas dans la modélisation de la distribution imposteur). Classiquement, le score normalisé S' s'obtient par la formule : $S' = \frac{(S - \mu_{imp})}{\sigma_{imp}}$, où μ_{imp} et σ_{imp} représentent la moyenne et la variance de la distribution imposteur respectivement.

4.2.7. Décision

La stratégie de décision est celle d'une tâche d'identification en milieu ouvert. Pour un segment, seul le locuteur ayant obtenu le score maximum est retenu (les autres sont rejetés). Une seule décision est prise (accepté/rejeté) pour ce locuteur et pour ce segment. Un seuil indépendant du genre est appliqué en vue du rejet ou de l'acceptation du locuteur. Le seuil a été appris empiriquement sur le corpus de développement *Dev*.

4.3. Présentation des spécificités de chaque système

Le LIA a présenté 5 systèmes différents dont 1 primaire. Ces cinq systèmes puisent leurs différences dans la façon de traiter la robustesse aux conditions de canal (Studio/Téléphone). Nous appliquons la même stratégie de décision et de normalisation des scores pour tous les systèmes.

4.3.1. Système 1 : *lia_svl_constrate-001_test*

Deux modèles du monde (Homme, Femme) sont appris sur les signaux d'apprentissage en condition Studio uniquement (utilisation d'EM). Les modèles de locuteurs sont adaptés avec leurs données d'apprentissage sans discrimination sur le canal.

4.3.2. Système 2 : *lia_svl_contraste-002_test*

Deux modèles du monde (Homme, Femme) sont appris sur les signaux d'apprentissage en condition Téléphone uniquement. Ces derniers sont obtenus grâce à l'adaptation MAP de leur modèle Studio correspondant. Les modèles de locuteurs sont adaptés à partir de l'ensemble des données disponible sans tenir compte du canal.

4.3.3. *lia_svl_primaire_test*

Le système primaire du LIA est construit à partir de la fusion des deux systèmes précédents. La fusion consiste en une moyenne arithmétique non pondérée des scores des systèmes *lia_svl_constrate-001_test* et *lia_svl_contraste-002_test*.

4.3.4. *lia_svl_contraste-003_test*

Les 4 modèles du monde des systèmes 1 et 2 sont utilisés pour ce système. Les locuteurs Homme et Femme sont divisés en sous-groupes, HommeTéléphone, HommeStudio, FemmeTéléphone, FemmeStudio. Cette division est effectuée en détectant le canal majoritaire pour chaque locuteur dans leurs signaux d'apprentissage. Ainsi le groupe HommeTéléphone comprend 48 locuteurs et HommeStudio 150. Les modèles de locuteurs sont ensuite adaptés avec leurs données d'apprentissage avec discrimination sur le canal.

4.3.5. *lia_svl_contraste-004_test*

Le dernier système du LIA se différencie par la stratégie employée lors de la paramétrisation des signaux et par le nombre de composantes des modèles qui est ici de 512, pour des problèmes de temps de calcul. Ce système utilise la technique dite du "Feature Mapping" pour l'étape de paramétrisation. Pour se faire, deux modèles, Homme et Femme, indépendants du canal ont été estimés, puis adaptés aux conditions Studio et Téléphone. Ces modèles sont utilisés pour projeter les paramètres dans un espace indépendant du canal. Il est à noter que la stratégie employée est une approximation du "Feature Mapping", puisque l'appartenance de la trame à tel ou tel canal est déterminée ici par la segmentation en macro-classes acous-

TAB. 1 – Taux d’erreurs de SES obtenus par le système du LIA sur les données d’évaluation *Eva*, exprimés sous la forme d’une F-Mesure.

<i>Eva</i>	
<i>Événement</i>	<i>F – Mesure</i>
musique	0.55
parole	0.99

tiques. L’état de l’art préconise le calcul de vraisemblances pour connaître le modèle auquel appartient cette trame.

5. Performances du système SES

Cette section présente les résultats obtenus par le système SES sur le corpus d’évaluation *Eva* de la campagne ESTER. La table 1 fournit les résultats obtenus par événement³. Comme décrit dans le descriptif du système SES, les résultats montrent que le système proposé n’a pas pour vocation de détecter la musique dans sa globalité (en présence ou non de parole) mais uniquement celle considérée comme de la non parole, en vue des processus situés en aval dans la chaîne des traitements (tâches SRL, TRS, ...). A l’opposé, le suivi de parole obtient pour sa part d’excellents résultats.

6. Performances du système SRL

Cette section présente les résultats obtenus par les systèmes SRL primaire et contrastif sur les ensembles de développement *Dev* et d’évaluation *Eva* de la campagne ESTER. La table 2 fournit les résultats globaux obtenus sur chacun des corpus⁴. Les tables 3 et 4 présentent les résultats individuels pour chacune des sources des corpus.

Ces résultats sont exprimés sous la forme de 4 taux d’erreurs :

- *MI* et *FA* sont relatifs aux erreurs de détection parole (*MI* - Missed Speaker - un segment locuteur, étiqueté non parole a été exclus automatiquement par le système) et de non parole (*FA* - False Alarm Speaker - un segment de non parole est étiqueté comme étant un locuteur potentiel) respectivement ;
- *Spk* correspondent aux erreurs de segmentation/regroupement en locuteur ;
- $All = MI + FA + Spk$;

Ces résultats montrent :

- des performances comparables entre les corpus *Dev* et *Eva* ;
- un meilleur comportement du système contrastif (comparé au système primaire) sur l’ensemble *Eva* ; aucune amélioration n’ayant pu être observée sur l’ensemble *Dev*, voire une légère baisse (de 16.9 à 17.2%) ;
- un comportement très variable des systèmes, quelque soit le corpus, selon la source observée. En effet, les taux d’erreurs en segmentation et regroupement (*Spk*) varient de 0 à 44.0%. Des résultats très variables sont observables pour une même radio : pour la radio RFI (*Eva*), les taux d’erreurs *Spk* fluctuent entre 0.5% et 31.4%. De même, pour une tranche horaire identique d’une même radio, des variations nettes sont observables : pour la ra-

dio RFI tranche 19h30-20h00 (*Eva*), les taux d’erreurs *Spk* varient de 19.7 à 31.4%.

Par ailleurs, une analyse du nombre de locuteurs détectés (voir la table 5 pour le détail de cette analyse) montre une sous-estimation flagrante comparé au nombre réel de locuteurs présents dans chacune des émissions. Ce déséquilibre est particulièrement notable pour les émissions d’une heure.

Discussion

L’évaluation ESTER a permis de confronter pour la première fois l’approche E-HMM à des émissions radiophoniques francophones. Seules des données issues des campagnes d’évaluation NIST des systèmes SRL avaient été utilisées jusqu’alors pour l’évaluation de ces performances ; ces données "américaines" de natures diverses (données téléphoniques, radiophoniques, et "meeting") avaient démontré l’efficacité de l’approche E-HMM (environ 8% d’erreur *Spk* sur les données radiophoniques lors de la campagne NIST RT’03).

La différence de performances entre les données "américaines" et "francophones" pourrait s’expliquer par différents facteurs :

- les durées des émissions à traiter sont très différentes d’un corpus à l’autre : les émissions américaines sont en moyenne d’une durée de 15 min contre 42 min pour les émissions francophones. Cette différence peut avoir un impact non négligeable sur le comportement du système, en particulier lors de l’application des techniques d’adaptation des modèles. En effet, le poids des modèles du locuteur L_0 ou des locuteurs majoritaires peut s’avérer différent (face aux nouveaux locuteurs détectés L_x) selon la quantité de données utilisée pour leur estimation ;
- si les corpus diffèrent dans la quantité de données, il en va de même du nombre de locuteurs à détecter, qui est sensiblement plus faible sur les émissions américaines. En particulier, les interventions sporadiques de locuteurs (moins de 30 secondes), difficiles à capter pour l’E-HMM, sont plus rares dans ces dernières que dans les émissions francophones.
- les émissions radiophoniques américaines, principalement des journaux, comportent moins de variabilité sur le plan acoustique que les émissions radiophoniques francophones. En effet, ces dernières comportent un grand nombre de reportages en extérieur, pouvant présenter une grande diversité de conditions acoustiques, notamment pour un même locuteur dans une même émission (voix chantées, fond musical, environnement très bruyé, ...), de nombreuses interventions de personnes inconnues (parole très spontanée), ... ;

L’analyse des résultats sur les corpus *Dev* et *Eva* a permis de montrer une déviance du comportement de l’approche E-HMM, accentuée par la durée croissante des émissions. Cette déviance conduit le système à proposer un "locuteur" majoritaire (parmi tous ceux détectés), englobant en fait différents locuteurs non détectés par le système (en particulier toutes les interventions de courtes durées) voire des portions de locuteurs déjà détectés par le système. Dans ce contexte, le nombre final de locuteurs détectés s’avère bien inférieur au nombre réel de locuteurs présents dans chacune des émissions.

Le système contrastif devait remédier en partie à ce problème en permettant à des locuteurs détectés par le système de récupérer les portions absorbées par ce locuteur "majoritaire" lors de la phase de resegmentation. Cet objectif est partiellement atteint au vu de l’amélioration observée sur le corpus *Eva*.

³Les références V1.3 ont été utilisées pour l’obtention des résultats sur le corpus *Eva*

⁴Les références V2.0 et V1.3 ont été utilisées pour l’obtention des résultats sur les corpus *Dev* et *Eva* respectivement

Des travaux en cours sur l'intégration de frontières "a priori" seraient une première piste pour pallier ce problème.

7. Performances du système SVL

Les résultats des différents systèmes sont donnés dans cette section sous forme de tableaux (tables 6 et 7) par fichier et par corpus (*Dev* et *Eva*). La segmentation de référence (fichiers *mdtm*) a été utilisée pour scorer l'expérience sur le corpus *Dev*. Lors de l'évaluation, les fichiers fournis par la tâche SRL sont été utilisés en entrée du système de Suivi du Locuteur. Le système *lia_contraste-003_test* n'a pu être évalué sur le *Dev*. Le système a été uniquement réglé sur un fichier d'une heure issu du corpus d'apprentissage (20030417_0800_0900_FRANCEINTER_DGA) sur lequel il obtenait une F-mesure de 0.81.

7.1. Discussion

De part l'hétérogénéité de ces résultats, quelques conclusions peuvent être tirées suivant les différentes stratégies utilisées pour chaque système :

- Les 5 systèmes ont des performances équivalentes sur le *Dev* et les systèmes ne semblent pas réagir de façon différente pour une source donnée.
- Le changement de comportement entre les corpus *Dev* et *Eva* est flagrant. Ceci est particulièrement dû à l'ajout des erreurs de segmentation en entrée des systèmes. Le système primaire du LIA est bien celui qui a donné les meilleures performances sur le set d'évaluation *Eva*.
- Le système 3, consistant à spécialiser les modèles de locuteurs par canal, a donné des résultats décevants. Ceci est probablement dû aux erreurs de segmentation en macro-classes (Studio et Téléphone). Les erreurs sont produites en amont et ne peuvent pas être "rattrapées" par le système de suivi.
- La fusion des systèmes 1 et 2 est très avantageuse. En regardant les réponses de chaque système sur les différentes sources, il semble que cette fusion permet à la fois de pallier les erreurs de segmentation en macro-classes acoustiques, mais aussi de répondre de façon optimale au canal en présence dans le fichier. Il faut rappeler que les signaux utilisés pour l'adaptation des modèles de locuteurs étaient multi-canaux ; il serait intéressant d'analyser plus en profondeur l'influence de tels modèles sur la robustesse d'un système de suivi du locuteur.
- Le système utilisant la technique du "Feature Mapping" obtient des résultats encourageants, proches des autres systèmes. Il faut rappeler que l'appartenance d'une trame à un modèle dépendant du canal est donnée par la segmentation en macro-classes acoustiques. Une stratégie différente, indépendante des erreurs de segmentation acoustique, devrait permettre d'améliorer les performances de ce système.
- Trois sources présentent des performances médiocres par rapport aux autres, par exemple 20041011_1300_1400_INTER_DGA. Après analyse, il s'avère que les erreurs de détection proviennent de la détection de genre utilisée en entrée. En effet, les tests effectués n'étant pas cross-gender, une erreur dans la détection de genre devient préjudiciable.
- Pour séparer les erreurs provenant du système de suivi du locuteur de celles du système de segmentation, il serait intéressant de refaire des expériences en utilisant la

segmentation de référence du corpus *Eva*.

7.2. Retour sur expériences et perspectives

Au cours de la campagne NIST 2000, le LIA a participé à une tâche de suivi du locuteur. Cette tâche ne présentait pas les mêmes difficultés tant au niveau du type de données (conversation téléphonique) qu'au volume de données à traiter (nombre de locuteurs à suivre limité à 1, quelques minutes de tests maximum). Les résultats obtenus lors de la campagne ESTER sur des données de type "broadcast news" sont par conséquent encourageants étant donné la difficulté de la tâche. Nous avons de ce fait identifié les enjeux et les voies de recherche à explorer pour ce type de tâches :

- Il serait intéressant d'analyser l'influence de données d'adaptation multi-canal sur les performances du système. Cette technique semble avoir amélioré la robustesse du système. On remarque qu'une fusion comme celle pratiquée est avantageuse et permet d'accepter une certaine marge d'erreur dans la segmentation macro-classes.
- La technique du feature mapping, utilisant la segmentation en macro-classes, ne permet pas de pallier les erreurs de segmentation. Le calcul de vraisemblance des trames sur des modèles dépendants du canal semble effectivement nécessaire pour appliquer cette technique.
- La quantité de données est un enjeu important et devient récurrent depuis les tâches Extended de NIST. Pourtant, ce n'est pas tant la quantité de données que ses variations dans l'apprentissage qu'il faudra étudier plus en détails. Ainsi, nous avons fixé une limite haute à la quantité de données disponible pour chaque locuteur (20 minutes comparées à une tâche NIST classique de 2,5 minutes). Nous pensons que les techniques d'adaptation ne sont pas adaptées à des durées d'apprentissage inter-locuteurs variables.

8. Conclusion

Ce papier décrit les différents systèmes présentés par le LIA lors de la première campagne d'évaluation ESTER pour les tâches de segmentation et de suivi d'événements uniquement. Cette campagne était organisée en janvier 2005. Tous ces systèmes ont été entièrement développés sur la plateforme ALIZE dédiée à la Reconnaissance Automatique du Locuteur.

Concernant la tâche SES ; le système du LIA obtient d'excellents résultats dès qu'il s'agit de suivre la parole. En revanche, des résultats plus médiocres sont obtenus pour le suivi de la musique. Cette contre-performance s'explique simplement par le fait que le système SES n'a pas été développé dans l'objectif de suivre la musique, mais plutôt de détecter les zones de non parole afin de les exclure en vue d'autres traitements. Détecter la musique en présence de parole n'était pas notre objectif premier.

Concernant la tâche SRL, le LIA participe depuis 2001 aux campagnes d'évaluation des systèmes SRL organisées par NIST. L'approche E-HMM sur laquelle repose le système du LIA a par conséquent été évaluée uniquement sur des corpus d'émissions radiophoniques américaines. La participation à la campagne ESTER a par conséquent permis d'évaluer pour la première fois le système SRL sur des corpus différents. Il est intéressant de constater des différences significatives dans les performances, fortement corrélées avec la nature des corpus (moyenne des durées des émissions très différentes par exemple entre les deux corpus). Cette comparaison montre que les per-

formances du système SRL lors de la campagne sont tout à fait satisfaisantes étant donné la "difficulté" du corpus, comparé au corpus américain. Une analyse du comportement du système dans ce nouveau contexte devrait permettre une progression significative des résultats.

Finalement la tâche SVL permet une nouvelle fois d'éprouver le système de détection du locuteur du LIA, après une participation à la campagne NIST SRE 2004. Malgré des résultats satisfaisants, le volume et le type de données disponibles pour l'évaluation ESTER, ainsi que la spécificité de la tâche de suivi du locuteur ont permis de mettre en évidence la nécessité d'adapter les stratégies de détection du locuteur à ce nouveau contexte.

9. References

- [1] J.-F. Bonastre, F. Wils, S. Meignier, Alize, a free toolkit for speaker recognition, in : Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2005), Philadelphia, USA, 2005.
- [2] S. Meignier, J.-F. Bonastre, C. Fredouille, T. Merlin, Evolutionary HMM for speaker tracking system, in : Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2000), Istanbul, Turquie, 2000, pp. 1177–1180.
- [3] S. Meignier, J.-F. Bonastre, S. Igounet, E-HMM approach for learning and adapting sound models for speaker indexing, in : 2001 : a Speaker Odyssey. The Speaker Recognition Workshop, Chania, Crete, 2001, pp. 175–180.
- [4] D. Moraru, S. Meignier, L. Besacier, J.-F. Bonastre, Y. Magrin-Chagnolleau, The ELISA consortium approaches in speaker segmentation during the NIST 2002 speaker recognition evaluation, in : Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2003), Vol. II, Hong Kong, 2003, pp. 89–92.
- [5] D. Moraru, S. Meignier, C. Fredouille, L. Besacier, J.-F. Bonastre, The ELISA consortium approaches in broadcast news speaker segmentation during the NIST 2003 rich transcription evaluation, in : Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2004), Montreal, Canada, 2004.
- [6] S. Meignier, D. Moraru, C. Fredouille, L. Besacier, J.-F. Bonastre, Benefits of prior acoustic segmentation for automatic speaker segmentation, in : Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2004), Montreal, Canada, 2004.
- [7] J. L. Gauvain, C. H. Lee, Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains, in : IEEE Transactions on Speech and Audio Processing, Vol. 22, 1994, pp. 291–298.

TAB. 2 – Taux d’erreur de SRL, exprimés en termes d’erreur de détection parole (Mi)/non parole (Fa), d’erreur de segmentation et regroupement en locuteur (Sp) et d’erreur totale (All) (en %), obtenus par les deux systèmes du LIA sur les données de développement Dev et d’évaluation Eva .

Système	<i>Dev</i>				<i>Eva</i>			
	<i>Mi</i>	<i>Fa</i>	<i>Spk</i>	<i>All</i>	<i>Mi</i>	<i>Fa</i>	<i>Spk</i>	<i>All</i>
Primaire	1.5	0.1	15.3	16.9	0.6	0.0	17.4	18.0
Contrastif	1.5	0.1	15.6	17.2	0.6	0.0	15.6	16.3

TAB. 3 – Taux d’erreur de SRL, exprimés en termes d’erreur de détection parole (Mi)/non parole (Fa), d’erreur de segmentation et regroupement en locuteur (Sp) et d’erreur totale (All) (en %), obtenus par les deux systèmes du LIA pour chacune des sources des données de développement Dev .

Sources	<i>Primaire</i>				<i>Secondaire</i>	
	<i>Mi</i>	<i>Fa</i>	<i>Spk</i>	<i>All</i>	<i>Spk</i>	<i>All</i>
20030418_0700_0800_FINT	0.2	0.6	12.2	13.0	12.1	12.9
20030418_0800_0900_FINT	0.3	0.2	18.9	19.3	18.9	19.3
20030418_1200_1300_FINF	3.0	0.1	15.8	19.0	16.0	19.2
20030418_1700_1800_FINF	3.1	0.0	28.9	32.1	29.0	32.1
20030508_1400_1500_RFI	0.9	0.0	15.7	16.7	15.7	16.7
20030509_1400_1500_RFI	0.9	0.0	13.9	14.7	14.0	14.9
20030717_0700_0715_RTM	1.9	0.0	44.0	45.9	43.6	45.6
20030717_1300_1320_RTM	1.0	0.0	1.8	2.8	7.0	8.0
20030717_2000_2020_RTM	1.3	0.0	0.1	1.4	0.1	1.4
20030717_2300_2315_RTM	1.3	0.0	21.0	22.3	20.0	21.3
20030719_0700_0715_RTM	2.3	0.0	0.4	2.7	0.4	2.7
20030719_1300_1320_RTM	1.5	0.0	12.6	14.1	13.1	14.5
20030719_2000_2015_RTM	2.1	0.2	0.1	2.4	0.0	2.4
20030719_2300_2310_RTM	3.6	0.0	0.0	3.6	0.0	3.6

TAB. 4 – Taux d’erreur de SRL, exprimés en termes d’erreur de détection parole (Mi)/non parole (Fa), d’erreur de segmentation et regroupement en locuteur (Spk) et d’erreur totale (All) (en %), obtenus par les deux systèmes du LIA pour chacune des données d’évaluation *Eva*.

Sources	Primaire				Secondaire	
	Mi	Fa	Spk	All	Spk	All
20041006_0700_0800_CLA	0.2	0.0	18.4	18.6	18.4	18.6
20041006_0800_0900_CUL	0.6	0.0	19.7	20.3	19.7	20.3
20041007_0800_0900_INT	0.2	0.0	14.3	14.5	8.2	8.5
20041008_1800_1830_INF	0.7	0.0	2.0	2.6	2.0	2.6
20041011_1300_1400_INT	0.2	0.0	16.4	16.7	12.5	12.7
20041012_1800_1830_INF	0.6	0.0	4.6	5.2	13.6	14.2
20041013_1700_1800_INF	1.3	0.0	28.2	29.5	20.9	22.2
20041025_1930_2000_RFI	1.1	0.0	19.7	20.8	21.8	22.9
20041026_1930_2000_RFI	0.5	0.0	31.4	32.0	34.1	34.7
20041027_1230_1300_RFI	0.1	0.0	8.3	8.4	7.4	7.6
20041124_1230_1300_RFI	0.1	0.0	0.5	0.6	0.6	0.7
20041217_1300_1322_RTM	0.6	0.0	10.2	10.9	10.2	10.9
20041218_1300_1314_RTM	0.6	0.0	15.1	15.7	10.8	11.4
20041219_1300_1314_RTM	0.3	0.0	43.8	44.1	22.6	23.0
20041220_1300_1314_RTM	0.5	0.0	24.7	25.2	24.7	25.2
20041221_1300_1321_RTM	1.0	0.0	23.8	24.8	23.8	24.8
20041222_1300_1320_RTM	2.1	0.0	33.7	37.7	33.7	35.7
20041223_1300_1318_RTM	2.2	0.0	9.7	11.9	9.6	11.8

TAB. 5 – Comparaison du nombre de locuteurs détectés par le système contrastif avec le nombre réel de ces locuteurs dans chacune des sources du corpus *Eva*.

Sources	Nb de locuteurs	
	réels	détectés
20041006_0700_0800_CLASSIQUE	27	21
20041006_0800_0900_CULTURE	17	12
20041007_0800_0900_INTER	38	20
20041008_1800_1830_INFO	23	17
20041011_1300_1400_INTER	26	19
20041012_1800_1830_INFO	30	19
20041013_1700_1800_INFO	60	28
20041025_1930_2000_RFI	18	15
20041026_1930_2000_RFI	31	15
20041027_1230_1300_RFI	16	14
20041124_1230_1300_RFI	20	14
20041217_1300_1322_RTM	11	10
20041218_1300_1314_RTM	5	4
20041219_1300_1314_RTM	5	7
20041220_1300_1314_RTM	6	5
20041221_1300_1321_RTM	8	4
20041222_1300_1320_RTM	12	12
20041223_1300_1318_RTM	8	6
Toutes confondues	361	242

TAB. 6 – Résultats des systèmes SVL du LIA sur ESTER DEV Phase II.

Source	C1	C2	C3	C4	P
20030418_1200_1300_FRANCEINFO_DGA	0.847	0.796	0.840	NC	0.841
20030418_1700_1800_FRANCEINFO_DGA	0.811	0.761	0.810	NC	0.804
20030418_0700_0800_FRANCEINTER_DGA	0.829	0.826	0.785	NC	0.843
20030418_0800_0900_FRANCEINTER_DGA	0.734	0.738	0.646	NC	0.743
20030508_1400_1500_RFI_ELDA	0.973	0.822	0.731	NC	0.919
20030509_1400_1500_RFI_ELDA	0.877	0.855	0.809	NC	0.878
20030717_0700_0715_RTM_ELDA	0.541	0.541	0.551	NC	0.541
20030717_1300_1320_RTM_ELDA	0.149	0.149	0.149	NC	0.149
20030717_2000_2020_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030717_2300_2315_RTM_ELDA	0.125	0.125	0.131	NC	0.125
20030719_0700_0715_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_1300_1320_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_2000_2015_RTM_ELDA	0.000	0.000	0.000	NC	0.000
20030719_2300_2310_RTM_ELDA	0.000	0.000	0.000	NC	0.000
F-mesure	0.629	0.597	0.572	NC	0.633

TAB. 7 – Résultats des systèmes SVL du LIA sur ESTER Test Phase II.

Source	C1	C2	C3	C4	P
20041008_1800_1830_INFO_DGA	0.582	0.743	0.251	0.319	0.644
20041012_1800_1830_INFO_DGA	0.398	0.295	0.390	0.189	0.393
20041013_1700_1800_INFO_DGA	0.511	0.400	0.458	0.346	0.512
20041007_0800_0900_INTER_DGA	0.622	0.813	0.575	0.581	0.804
20041011_1300_1400_INTER_DGA	0.332	0.332	0.300	0.330	0.332
20041006_0700_0800_CLASSIQUE	0.000	0.000	0.000	0.000	0.000
20041006_0800_0900_CULTURE	0.000	0.000	0.000	0.000	.000
20041025_1930_2000_RFI_ELDA	0.730	0.251	0.113	0.505	0.826
20041026_1930_2000_RFI_ELDA	0.546	0.405	0.000	0.493	0.556
20041027_1230_1300_RFI_ELDA	0.000	0.000	0.000	0.000	0.000
20041124_1230_1300_RFI_ELDA	0.000	0.000	0.000	0.000	0.000
20041217_1300_1322_RTM_ELDA	0.749	0.748	0.786	0.839	0.748
20041218_1300_1314_RTM_ELDA	0.000	0.000	1.000	0.000	0.000
20041219_1300_1314_RTM_ELDA	0.000	0.000	0.000	0.000	0.000
20041220_1300_1314_RTM_ELDA	0.799	0.839	0.822	0.984	0.841
20041221_1300_1321_RTM_ELDA	0.747	0.778	0.322	0.804	0.747
20041222_1300_1320_RTM_ELDA	0.711	0.673	0.593	0.671	0.712
20041223_1300_1318_RTM_ELDA	0.681	0.646	0.000	0.588	0.681
F-mesure	0.605	0.585	0.413	0.507	0.659