Segmentation en macro-classes acoustiques d'émissions radiophoniques dans le cadre d'ESTER

Corinne Fredouille, Driss Matrouf, Georges Linarès, Pascal Nocera

Segmentation en macro-classes acoustiques d'émissions radiophoniques dans le cadre d'ESTER

Corinne Fredouille, Driss Matrouf, Georges Linares, Pascal Nocera

LIA-Avignon – BP 1228 – 84911 Avignon Cedex 9 - France
Mél: (corinne.fredouille, driss.matrouf, georges.linares, pascal.nocera)@lia.univ-avignon.fr

ABSTRACT

This paper is dedicated to the study of automatic acoustic segmentation systems, often required for broadcast news show processing. This study is carried in the framework of the phase I of the French evaluation campaign, called ESTER. It aims at evaluating the behavior of an acoustic segmentation system, especially tuned for American broadcast news show processing, when it is applied to French ones. Furthermore, different implementations of this system are evaluated, each of them taking into account French data. The goal of these implementations is to demonstrate which kind of French data have to be used and how to integrate them in the acoustic segmentation system. Results obtained through various experiments highlight interesting issues.

1. INTRODUCTION

La segmentation de signaux audio en macro-classes acoustiques relatives à différents environnements est devenue une phase importante voire nécessaire à tout système traitant des émissions radiophoniques en vue d'améliorer leur performance. Introduite initialement comme phase préliminaire d'un système de transcription automatique, elle avait pour objectif de rejeter les portions de non parole (silence, musique, …). Par la suite, la classification fournie par les systèmes de segmentation s'est affinée pour permettre l'adaptation des modèles acoustiques à des environnements particuliers tels que la parole téléphonique ou la parole sur fond musical ou, plus simplement, au genre des locuteurs [1][2][3]. Plus récemment, cette phase de segmentation s'est révélée intéressante pour d'autres tâches liées au traitement des émissions radiophoniques telles que la segmentation automatique en locuteurs [4].

L'étude présentée dans ce papier a été réalisée dans le cadre de la phase I de la campagne ESTER (Evaluation des Systèmes de Transcription Enrichie d'émissions Radiophoniques) [5]. Cette campagne, organisée dans le cadre du projet EVALDA/Technolangue (financé par le Ministère français de la Recherche) a pour objectif de favoriser le développement de la recherche dans le domaine du traitement de la parole en langue française et notamment de promouvoir une dynamique de l'évaluation. Pour répondre à ces objectifs, une diffusion large des informations et ressources (corpus transcrits, outils, …) relatives à ces évaluations a débuté en septembre 2003 auprès des participants.

Cette étude préliminaire est dédiée à la segmentation en macro-classes acoustiques. Elle a pour objectif d'évaluer la portabilité d'un système de segmentation, développé exclusivement pour le traitement d'émissions radiophoniques américaines, dans le cadre d'ESTER. Par ailleurs, différentes variantes de ce système sont étudiées afin d'évaluer quelles données de la campagne d'ESTER utiliser et quelles méthodes choisir pour les intégrer.

La section 2 est dédiée à la description du système de segmentation en macro-classes acoustiques. Les différentes variantes du système sont présentées en section 3. Les résultats obtenus sont présentés et commentés en section 4. Finalement, un résumé de cette étude est proposé en section 5, suivi de quelques perspectives sur la continuité de ce travail.

2. SYSTÈME DE SEGMENTATION EN MACRO-CLASSES ACOUSTIQUES

Le système de segmentation de base repose sur une approche hiérarchique, décrite en section 2.2. Il peut fournir une segmentation en trois niveaux : parole/non parole, parole/parole sur fond musical/parole téléphonique, détection du genre des locuteurs.

2.1. 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 25ms), de la log-énergie normalisée ainsi que des dérivées premières et secondes. Aucune normalisation des paramètres n'est appliquée. Aucun traitement particulier n'est utilisé pour la parole téléphonique.

2.2. Approche hiérarchique

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

- Durant la première étape, une segmentation parole/non parole est appliquée sur le signal (une émission complète). Deux modèles, « MixS » et « NS » sont utilisés lors de cette segmentation, représentant respectivement les conditions *parole* et *non parole*. Ce processus de segmentation repose sur une recherche du meilleur modèle au niveau de la trame, suivie de l'application de règles morphologiques pour l'aggrégation des trames suivant les labels *parole* et *non parole*. Par exemple, un segment contenant une majorité de trames *non parole* est considéré comme tel si sa durée excède 0,4s.
- La seconde étape a pour objectif de segmenter les portions *parole*, détectées préalablement, suivant trois nouvelles étiquettes : *parole*, *parole téléphonique* et *parole sur fond musical*. Ces étiquettes sont représentées respectivement par trois modèles indépendants du genre: « S », « T » et « MS ». Ce processus de segmentation repose, ici, sur un décodage Viterbi, appliqué sur un HMM ergodique à trois états (les modèles « S », « T » et « MS »).
- Finalement, la dernière étape se focalise sur une détection du genre, durant laquelle chaque segment issu de la phase précédente est étiqueté soit homme soit femme. Cette détection s'appuie sur le même principe de décodage que l'étape 2, associé à des modèles dépendants du genre et de la classe acoustique du segment traité. Ainsi, les modèles

« GS-Fe » et « GS-Ma » représentent respectivement des segments de parole produits par des femmes et des hommes, « GT-Fe » et « GT-Ma » des segments de parole téléphonique femmes et hommes et « GSM-Fe » et « GSM-Ma » des segments de parole sur fond musical femmes et hommes.

Tous les modèles utilisés lors des différentes phases de segmentation sont des mixtures de gaussiennes (GMM) caractérisées par des matrices de covariance diagonales. Excepté les modèles « NS » et « MixS » composés respectivement de 1 et 512 gaussiennes, tous les autres modèles sont caractérisés par 1024 gaussiennes. Ces nombres de composantes ont été fixés empiriquement. Les outils d'apprentissage des modèles ainsi que le décodeur Viterbi sont issus du toolkit HTK [6].

3. PROTOCOLES EXPERIMENTAUX

Les expériences présentées dans ce papier ont pour objectif d'évaluer les performances du système de segmentation dans le cadre d'ESTER. Cette évaluation porte sur la détection parole/non parole, la segmentation parole/parole téléphonique/parole sur fond musical ainsi que sur la détection en genre des locuteurs. Quatre variantes du système de segmentation sont présentées et testées dans ce papier.

3.1. Bases de données

Plusieurs ensembles de données acoustiques ont été utilisés dans le cadre de ce travail :

- Le sous-ensemble *BN–Seg,* extrait du corpus des *Broadcast News 96* développé spécifiquement pour les campagnes d'évaluation américaines des systèmes automatiques de transcription (projet HUB-4) d'émissions radiophoniques et télévisées américaines. Le sous-corpus *BN–Seg* est constitué d'ensembles de signaux d'environ 2h chacun, représentant différents environnements acoustiques : parole, parole téléphonique, parole sur fond musical, chacun décliné en parole homme, parole femme, parole mixte.
- Les corpus *FI+RFI–Seg* et *FI–Seg* extraits du corpus ESTER-Train-Phase I. Ce dernier corpus, est distribué dans le cadre de la phase I de la campagne ESTER. Dédié à l'apprentissage (ensemble Train), il est composé de 19h40 d'émissions de France Inter (FI) et de 11h d'émissions de Radio France International (RFI). D'une manière similaire au corpus *BN–Seg,* nous avons sélectionné des sous-ensembles de signaux, d'une durée d'environ 2h chacun, parmi les émissions de FI et RFI pour le corpus *FI+RFI–Seg* et uniquement de FI pour le corpus *FI–Seg*.
- Le corpus *Dev* fourni dans le cadre d'ESTER pour le développement des systèmes de la phase I. Ce corpus est composé de signaux d'une durée totale de 4h40, provenant d'émissions de FI (*Dev/FI* 2h40) et de RFI (*Dev/RFI* 2h).

3.2. Variantes du système de segmentation

Le système de segmentation en macro-classes acoustiques a été initialement développé comme phase préliminaire du système de segmentation en locuteurs du LIA. Testés lors de la campagne d'évaluation américaine de transcription enrichie, NIST/RT'03 [7], les deux systèmes combinés ont été appliqués sur des émissions radiophoniques américaines [4]. Dans ce contexte, l'ensemble des modèles GMM utilisés par le système de segmentation en macro-classes acoustiques a été appris sur le corpus *BN–Seg*.

Dans le cadre de la campagne d'évaluation ESTER, un système de segmentation en macro-classes acoustiques peut être nécessaire pour répondre à différentes tâches de la campagne telles que, par exemple, la tâche de suivi d'événements sonores (SES), de transcription orthographique (TTR) ou de segmentation/regroupement en locuteurs (SRL).

L'approche hiérarchique a été choisie dans ce contexte pour fournir une segmentation acoustique similaire à celle obtenue pour le traitement des émissions radiophoniques américaines. La seule différence porte principalement sur les données utilisées (décrites en section 3.1) pour l'estimation des GMM du système de segmentation (les GMM utilisés pour la détection parole/ non parole ne sont pas concernés ici et restent appris sur le corpus *BN-Seg*.). Différentes variantes sont proposées et testées dans ce papier :

- *BN* : Apprentissage des GMM sur le corpus *BN–Seg,* (système initial) ;
- *EST.* : Apprentissage des GMM sur le sous-ensemble *FI+RFI–Seg* ;
- *BN/FI* : Adaptation des GMM, appris sur le corpus *BN–Seg*, avec le sous ensemble *FI–Seg* ;
- *BN/FI+RFI* : Adaptation des GMM, appris sur le corpus *BN–Seg*, avec le sous ensemble *FI+RFI–Seg*.

L'adaptation des GMM (moyenne et variance) mentionnée ci-dessus repose sur l'application successive des méthodes MLLR et MAP. Les outils d'adaptation sont issus du toolkit HTK [6].

3.3. Protocole d'évaluation

Les différentes variantes du système de segmentation (incluant les données d'ESTER), ainsi que la version initiale (version américaine) sont testées sur le corpus *Dev* ainsi que sur les corpus *Dev/FI* et *Dev/RFI* en vue d'étudier le comportement des systèmes sur chacune des sources radiophoniques. La métrique d'évaluation repose sur une comparaison trame à trame des étiquettes données par le système automatique (Hyp) et celles de référence (Ref). On calcule ainsi un taux de classification correcte (resp. incorrecte) par le rapport du nombre de trames S_{Ref} étiquetées S_{Hyp} (resp. X_{Hyp}) sur le nombre total de trames S_{Ref}.

4. RÉSULTATS ET DISCUSSION

4.1. Classification S, MS, T, NS

Les tables 1A, 1B, 1C présentent les résultats du système de segmentation et ses variantes suivant les classes : *parole* (S), *parole sur fond musical* (MS) et *parole téléphonique* (T) sur les corpus *Dev*, *Dev/FI* et *Dev/RFI* respectivement.

- Pour la classe *parole*, la variante *BN/FI* obtient le meilleur taux de classification correcte sur *Dev* (90.7%), devant la variante *EST* (86%). L'utilisation des données conjointe de FI et de RFI pour adapter les modèles GMM dégrade les performances, comparées à l'utilisation de FI seule. Néanmoins, les résultats individuels sur chacune des sources révèlent des comportements différents. Les taux sur *Dev/FI* sont meilleurs sur des données mixtes (93.2% pour *BN/FI+RFI*) avec une préférence pour des données uniquement françaises (95,4% pour *EST.*). Les taux sur *Dev/RFI* se dégradent singulièrement dès lors que des données RFI sont utilisées (56.6 et 67.0% pour *BN/FI+RFI* et *EST.* contre 73 et 89.4% pour *BN* et *BN/FI*). Dans tous les cas, les erreurs se reportent en grande majorité sur la classe *parole sur fond musical*.
- Pour la classe *parole téléphonique*, les taux de classification correcte sont quasiment constants quelle que soit la variante utilisée. Une légère amélioration de ces taux

est observée sur *Dev/FI;* ~98% contre 96 à 97% sur *Dev/RFI*. Il est à noter que *Dev/RFI* comporte 2 fois plus de signaux téléphoniques que *Dev/FI* (31 contre 11.7%) et comprend de la parole téléphonique sur fond musical contrairement à *Dev/FI,* source de dégradation potentielle.

- Pour la classe *parole sur fond musical*, les meilleurs taux, sur le corpus *Dev*, sont obtenus par les variantes *BN* et *EST*. Les résultats obtenus sur chacune des sources montrent une dégradation des performances sur *Dev/FI* dès lors que des données RFI sont utilisées pour l'adaptation et inversement une perte sur *Dev/RFI* si des données FI sont utilisées seules. Il est intéressant de noter que la perte importante observée sur *Dev/FI* n'est pas répercutée sur la variante *EST*. qui tient compte pourtant des deux sources radiophoniques.
- Pour la classe *non parole*, un taux de classification correcte de 90.9% est atteint sur le corpus *Dev,* le système se comportant mieux sur *Dev/RFI* (93.3%) que sur *Dev/FI* (seulement 78.9%). Il est à noter que le système de segmentation est développé pour détecter les portions de non parole dont la durée est supérieure à 0,4s. Hors, une grande majorité des longs silences du corpus ESTER n'ont pas été étiquetés en tant que tel et sont donc considérés comme de la parole, entraînant des erreurs de classification.

L'analyse de ces résultats montre que la détection de la parole téléphonique ne dépend pas des données utilisées pour l'apprentissage des GMM. Seule une bonne modélisation de l'espace fréquentiel (prise en compte de la bande passante téléphonique) est nécessaire.

Par ailleurs, l'utilisation conjointe des données des deux sources radiophoniques FI et RFI, en apprentissage (variante *EST*.) ou en adaptation (variante *BN/FI+RFI*) facilite la segmentation du corpus *Dev/FI* pour la classe *parole* mais perturbe fortement celle du corpus *Dev/RFI*. La particularité des données RFI, qui réside en une grande variabilité dans les accents des locuteurs, peut être une première explication à ce comportement atypique.

Concernant la classe *parole sur fond musical*, l'adaptation des GMM semble problématique selon les données d'adaptation utilisées. Si la quantité de données est suffisante, il semble préférable d'apprendre directement les modèles acoustiques.

Finalement, la variante *BN* obtient des résultats très honorables, voire dans certains cas, équivalents aux autres variantes intégrant, quant à elles, des données françaises. Ce résultat tend à montrer qu'un système de segmentation indépendant de la langue peut être utilisé dès lors que peu de signaux sont disponibles pour une langue cible.

4.2. Détection du genre des locuteurs

Les tables 2A, 2B, 2C présentent les résultats de la détection du genre des locuteurs (classe *homme* ou *femme*) sur les corpus *Dev*, *Dev/FI* et *Dev/RFI* respectivement.

- Pour la classe *homme*, les taux de classification varient légèrement d'une variante à l'autre, le meilleur taux étant obtenu avec la variante *BN/FI* (93.8%), suivie de la variante *BN* (93.5%). Un comportement plutôt similaire est observé sur chacune des sources malgré des taux plus faibles sur *Dev/RFI*. Les erreurs de classification se répartissant de manière quasi égale entre la classe *femme* et la classe *non parole*.
- Pour la classe *femme*, les taux sont plus faibles avec une dégradation très nette sur la variante *BN* (de 85.6 pour le meilleur taux à 65.3% pour *BN*). Cette dégradation est causée par la source RFI qui enregistre une perte de plus de 50% en absolu sur cette même variante (43.7%), comparé à son meilleur résultat (80.9% avec *BN/FI+RFI*). Les erreurs de classification étant principalement dues à une confusion femme/homme.

D'après ces résultats, la détection des femmes semble plus problématique que celle des hommes. Les taux de détection correcte sont plus faibles et l'utilisation de modèles appris sur une autre langue dégrade les performances. Ce phénomène est accentué sur la source RFI, conduisant à des écarts de 5 à 10% avec les taux obtenus sur la source FI.

5. CONCLUSION

Une première étude sur la segmentation automatique en macro-classes acoustiques, menée dans le cadre de la phase I de la campagne ESTER, est proposée dans ce papier. Elle a pour objectif d'observer, d'une part, le comportement d'un système appris exclusivement sur des données américaines pour segmenter des émissions radiophoniques françaises, et d'autre part, d'étudier des variantes de ce système, tenant compte, quant à elles, de données françaises. Les résultats des tests réalisés sur le corpus de développement de la phase I d'ESTER ont montré que le système américain obtient des résultats de segmentation tout à fait satisfaisants sur le français et surtout encourageants s'ils s'avèrent généralisables à des langues étrangères pour lesquelles très peu de données audio sont disponibles. Concernant les différentes variantes et aux vues des résultats, il semble préférable d'estimer directement des modèles (si la quantité de données est suffisante) pour le segmenteur, dès lors que des données issues de plusieurs sources radiophoniques sont utilisées. En effet, le comportement des variantes du segmenteur basées sur l'adaptation des modèles américains semble plus instable suivant les sources à traiter.

En fait, une étude approfondie de la source RFI et de son impact sur l'estimation des modèles du segmenteur sera nécessaire pour mieux comprendre les raisons de cette instabilité. En outre, la classe *parole* comprend actuellement la parole propre (studio) et celle dégradée (extérieur). L'étude d'une segmentation plus fine pourrait permettre d'améliorer les performances du segmenteur à ce niveau. Un effort est également à fournir pour la détection homme/femme.

BIBLIOGRAPHIE

[1] P.C. Woodland, "The development of the HTK Broadcast News transcription system: An overview", *Speech Communication*, Vol. 37, pp. 291-299, 2002.

[2] J.L. Gauvain, L. Lamel, and G. Adda. "The LIMSI Broadcast News Transcription System". *Speech Communication*, 37(1-2):89-108, 2002.

[3] T. Hain, and P.C. Woodland, "Segmentation and Classification of Broadcast News audio", *ICSLP'98*, Sydney, Australia.

[4] S. Meignier, D. Moraru, C. Fredouille, L. Besacier, and J.-F. Bonastre, "Benefit of prior acoustic segmentation for speaker segmentation systems". *Papier accepté à ICASSP'04,* Montréal, Canada.

[5] « Campagne ESTER: Evaluation des Systèmes de Transcription Enrichie d'émissions radiophoniques », Plan d'évaluation de la phase I. http://www.afcp-parole.org/ester/

[6] « HTK toolkit », http://htk.eng.cam.ac.uk/

[7] "NIST/RT'03", http://www.nist.gov/speech/tests/rt/rt2003/spring/index.htm

TABLES 1A, 1B, 1C: Résultats du système de segmentation et ses variantes (*BN*, *BN–FI*, *BN–FI+RFI* et *EST*) en classes acoustiques: *parole* (S), *parole téléphonique* (T), *parole sur fond musical* (MS) et *non parole* (NS). Comparaison des résultats suivant les corpus *Dev*, *Dev/FI* et *Dev/RFI*. La quantité de données associée à chacune des classes, exprimée en % de trames (% Tr.) et calculée sur l'ensemble du corpus, est fournie pour indication.

TABLE 1A: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev* (4h40 d'émissions).

Hyp	**Parole (S) (en %)**				**Téléphone (T) (en %)**				**Parole+Musique (MS) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
S (68.7)	**82.0**	**90.7**	**81.2**	**86.1**	0.8	0.7	0.8	0.7	13.5	4.9	14.3	9.5	3.7
T (20.1)	0.6	1.2	0.4	1.0	**97.3**	**96.8**	**97.4**	**96.9**	0.1	0.0	0.2	0.1	2.0
MS (8.3)	11.2	27.2	18.6	10.8	0.0	0.0	0.0	0.0	**79.8**	**63.8**	**72.4**	**80.2**	9.0
NS (2.9)	0.7	1.4	2.1	2.5	0.2	0.01	0.0	0.0	8.2	7.7	7.0	6.6	**90.9**

TABLE 1B: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev/FI* (2h40 d'émissions).

Hyp	**Parole (S) (en %)**				**Téléphone (T) (en %)**				**Parole+Musique (MS) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
S (82.0)	**86.4**	**91.4**	**93.2**	**95.4**	0.0	0.0	0.0	0.0	10.9	5.9	4.1	1.9	2.7
T (11.7)	0.5	0.7	0.6	1.0	**98.3**	**98.1**	**98.2**	**97.8**	0.0	0.0	0.0	0.0	1.2
MS (5.4)	16.2	10.8	35.2	12.4	0.0	0.0	0.0	0.0	**78.3**	**83.7**	**59.3**	**82.1**	5.5
NS (0.9)	2.5	3.0	11.0	12.8	0.9	0.0	0.0	0.0	17.7	18.1	10.1	8.3	**78.9**

TABLE 1C: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev/RFI* (2h d'émissions).

Hyp	**Parole (S) (en %)**				**Téléphone (T) (en %)**				**Parole+Musique (MS) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
S (51.4)	**73.0**	**89.4**	**56.6**	**67.0**	2.2	2.1	2.1	2.1	18.9	2.6	35.4	25.0	5.9
T (31.0)	0.6	1.5	0.3	1.1	**96.9**	**96.1**	**97.0**	**96.5**	0.1	0.0	0.3	0.0	2.4
MS (12)	8.3	36.7	8.9	10.0	0.0	0.0	0.0	0.0	**80.7**	**52.3**	**80.1**	**79.0**	11.0
NS (5.6)	0.4	1.0	0.3	0.5	0.0	0.0	0.0	0.0	6.3	5.7	6.4	6.2	**93.3**

TABLES 2A, 2B, 2C: Résultats du système de segmentation et ses variantes (*BN, BN-FI, BN-FI+RFI* et *EST*) en classes acoustiques: *homme* (H) et *femme* (F). Comparaison des résultats suivant les corpus *Dev*, *Dev/FI* et *Dev/RFI*. La quantité de données associée à chacune des classes, exprimée en % de trames (% Tr.) et calculée sur l'ensemble du corpus, est fournie pour indication.

TABLE 2A: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev* (4h40 d'émissions).

Hyp	**Homme (H) (en %)**				**Femme (F) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
H (73.9)	**93.5**	**93.8**	**92.8**	**93.1**	2.5	2.2	3.2	2.9	4.0
F (23.1)	31.6	12.9	11.2	13.7	**65.3**	**83.9**	**85.6**	**83.2**	3.1

TABLE 2B: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev/FI* (2h40 d'émissions).

Hyp	**Homme (H) (en %)**				**Femme (F) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
H (73.9)	**94.1**	**95.5**	**93.6**	**94.7**	3.2	1.8	3.7	2.6	2.7
F (23.1)	13.5	8.9	7.6	11.8	**83.7**	**88.3**	**89.6**	**85.4**	2.8

TABLE 2C: Taux de classification correcte (**en %**) et incorrecte (en %) sur le corpus ESTER- Phase I *Dev/RFI* (2h d'émissions).

Hyp	**Homme (H) (en %)**				**Femme (F) (en %)**				**Non par. (NS) (%)**
Ref (% Tr.)	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*	*BN/ FI*	*BN/ FI+RFI*	*EST.*	*BN*
H (73.9)	**92.5**	**91.4**	**91.6**	**91.0**	1.5	2.6	2.4	3.0	6.0
F (23.1)	52.6	17.5	15.4	15.7	**43.7**	**78.8**	**80.9**	**80.6**	3.7