euh, rire et bruits en parole spontanée : application à l'alignement forcé

Brigitte Bigi, Christine Meunier

euh, rire et bruits en parole spontanée : application à l'alignement forcé

Brigitte Bigi, Christine Meunier
Laboratoire Parole et Langage, CNRS, Aix-Marseille Université
13100 Aix-en-Provence, France
brigitte.bigi@lpl-aix.fr, christine.meunier@lpl-aix.fr

RESUME

Contrairement à la parole contrôlée, dans laquelle les intentions du locuteur sont très restreintes, la parole spontanée fait référence à une activité plus libre mais aussi plus riche de facteurs caractéristiques de l'interaction langagière . A ce titre, de nombreux phénomènes apparaissent comme les hésitations, les mots tronqués, les réductions phonétiques, etc. Nous proposons dans un premier temps un recensement de 3 événements paralinguistiques ("euh", rire, bruit), dans différents corpus spontanés : débat politique, narration, dialogue orienté tâche, dialogue informel avec consigne et dialogue informel sans consigne. Bien que ces événements soient fréquemment produits par les locuteurs, nous observons des différences significatives selon les corpus. A titre applicatif, nous montrons que les résultats de l'alignement forcé peuvent être nettement améliorés lorsque le système dispose d'un modèle acoustique qui inclut ces événements.

ABSTRACT

Filled pause, laughter and noise in spontaneous speech: application to forced-alignment.
Contrariwise to controlled speech, for which speaker's intention are very limited, spontaneous speech refers to a freer but also richer activity that is characteristic of language interaction., . Many phenomena appear like hesitations, truncated words, phonetic reductions, etc. In this paper, we first propose a frequency survey of 3 paralinguistic events ("uh", laughter, noise), in different spontaneous corpora: political debate, interviews, task-oriented dialog, informal dialog with instructions and informal dialog without instructions. Even if these events are frequently produced by the speakers, we observe significant differences according to the corpora. For illustrative purposes, we finally show that the results of forced-alignment are significantly improved when the acoustic model of the system includes these events.

MOTS-CLES : parole spontanée, euh, rire, bruit, alignement forcé.
KEYWORDS: spontaneous speech, filled pause, laughter, noise, forced-alignment.

1 Introduction

La tâche d'alignement forcé consiste à déterminer automatiquement la localisation temporelle des phonèmes d'un fichier audio. Tandis que l'alignement de la parole lue a été largement abordée, les études relatives à la segmentation de la parole spontanée restent peu nombreuses. Les productions de parole extraites de situations naturelles et occasionnelles sont caractérisées par un débit de parole rapide mais aussi irrégulier, des troncatures de mots, des réductions de phonèmes (Johnson, 2004), etc. Le discours spontané est en effet produit dans une situation communicative dynamique impliquant des routines linguistiques et des contraintes qui conduisent à une réorganisation de la production sonore, puis à des variations massives. Ces caractéristiques entraînent de grandes difficultés lorsque le flux vocal doit être annoté en unités phonétiques discrètes.

Les différents types de paroles spontanées peuvent fournir divers degrés de réduction, selon que la production est plus ou moins contrôlée. Une difficulté importante pour les outils d'alignement automatique réside dans le fait que la réduction n'est pas systématiquement une suppression discrète de phonèmes. Plusieurs études (Adda-Decker et al., 2013, Meunier, 2013) ont montré que, très souvent, la réduction entraîne une coalescence des phonèmes (plusieurs phonèmes sont fusionnés en un segment). Ces instances sont assez fréquentes et ne sont généralement pas perçues par les transcripteurs. De plus, la parole spontanée est caractérisée par plusieurs éléments qui n'apparaissent pas en condition contrôlée. En particulier rires, toux, bruits de bouche, etc. apparaissent fréquemment en conversation. Certains travaux (Ogden, 2001) indiquent que les *clics*, par exemple, sont utilisés de manière linguistique pour structurer le discours oral. Ces éléments n'appartiennent pas aux inventaires phonologiques, cependant, ils sont très présents dans les conversations par exemple et les outils automatiques doivent les identifier afin de fournir un alignement phonétique correct.

L'étude que nous proposons dans cet article s'inscrit dans le cadre de la linguistique de corpus. A partir des exemples contenus dans des données réelles, nous recensons différents événements de la parole spontanée. Nous avons ainsi sélectionné un ensemble de 5 corpus, relativement homogènes dans leur forme, de sorte que nous puissions observer et comparer les variations de fréquences des tokens. Nous nous focalisons sur 3 événements bien particuliers : le "euh", le rire et le "bruit". Dans un deuxième temps, nous montrerons où se situent ces événements relativement au reste de la parole : isolés entre deux silences, en début de segment, en fin de segment ou au sein d'un segment de parole. Nous avons ensuite évalué l'impact de la prise en compte de ces derniers dans la tâche d'alignement automatique. Pour ce faire, nous avons comparé les résultats d'un même système selon qu'il utilise un modèle acoustique qui inclut soit une représentation prototypique, soit un modèle spécifiquement appris pour chacun des trois événements euh, rire et bruit.

2 Corpus collectés

2.1 Origine et description des corpus

Pour cette étude, nous avons réuni et sélectionné les corpus décrits en table 1, afin qu'ils soient les plus homogènes possible. La première colonne indique le nom couramment donné au corpus, la deuxième le type d'enregistrement, la troisième la durée de parole (c'est-à-dire qu'elle n'inclut pas les silences), enfin, la quatrième colonne raporte le nombre de locuteurs ainsi que le style de parole qui est détaillé entre parenthèses.

Nom	**Enreg.**	**Durée de parole**	**Loc.**	**Style de parole**
Europe	audio	33 min	6	Débat politique à la radio
Typaloc	audio	39 min	4	Conversation (interview)
AixMapTask	audio-vidéo	163 min	10	Conversation (orientée tâche)
CID	audio-vidéo	7h30min	16	Conversation (dialogue informel)

Cheese	audio-vidéo	63 min	8	Conversation (dialogue informel)

TABLE 1 : Description des corpus

Dans la mesure où notre étude concerne spécifiquement certains événements paralinguistiques, il était important qu'aucun autre paramètre ne puisse influencer voire biaiser les résultats. Tous les corpus qui ont été sélectionnés ont été enregistrés en chambre sourde, avec un micro par locuteur. Chaque signal audio a été automatiquement segmenté en Unités Inter-Pausales (IPUs), i.e. des segments de production sonore alignés sur le signal. Ces segments sont entourés de silences dont la durée dépasse 200ms. Les frontières des IPUs ont été vérifiées manuellement pour chacun des corpus. Une transcription orthographique enrichie a ensuite été réalisée au sein des IPUs, en suivant la convention de transcription spécifique au logiciel SPPAS[1] (Bigi, 2015). Ces transcriptions comprennent : les pauses pleines (euh), les rires (@), les bruits (*), les pauses courtes (+), les disfluences (répétitions, mots tronqués, …), les prononciations inhabituelles ainsi que les élisions inhabituelles. Compte-tenu des conditions d'enregistrement, dans les corpus de la table 1, les bruits se limitent à des productions du locuteur, à savoir des souffles, respirations, toux, etc. Enfin, dans *Europe* et *CID*, tous les euh, rires et bruits ont été alignés manuellement sur le signal.

Le corpus *Europe* (Portes, 2004) est un débat politique enregistré sur une station de radio essentiellement dédiée à l'information. Ce débat implique quatre invités interrogés par deux journalistes à propos de l'Union Européenne et plus particulièrement de la question délicate de ses frontières. Le corpus *TYPALOC* (Meunier *et al.*, 2016) original se compose de plusieurs corpus de lectures (mots et textes) et de parole spontanée (interviews), produites par des locuteurs en bonne santé et des locuteurs affectés par une dysarthrie. Pour cette étude, nous avons conservé uniquement les interviews (8-17 min) de locuteurs en bonne santé. La condition audio-visuelle du corpus *AixMapTask*[2] se compose d'enregistrements audio et vidéo de dialogues orientés tâche (Gorish *et al.*, 2014). Le protocole expérimental suit les règles standards d'une Map Task : les participants sont autorisés à dire tout ce qu'ils veulent dans le but d'accomplir la tâche qui leur incombe. Dans ce corpus, les participants sont assis face-à-face avec leurs cartes respectives posées sur des pupitres. Le *Corpus of Interactional Data - CID*[3] (Bertrand *et al.*, 2008) est constitué de dialogues d'une heure chacun, enregistrés en audio et en vidéo. L'un des deux sujets de conversation a été suggéré aux participants : des conflits dans leur environnement professionnel (5 dialogues) ou des situations insolites auxquelles ils ont été confrontés (3 dialogues). *Cheese* (Priego-Valverde & Bigi, 2016) est également un enregistrement audio-vidéo de dialogues (~ 15 min) impliquant deux participants. Il a été demandé à chacun de lire une blague, imposée par l'expérimentateur, puis de discuter librement. La partie relative à la lecture a été retirée du corpus pour la présente étude.

2.2 Distributions des tokens

Chacun de ces corpus a été normalisé avec l'outil "Text Normalization" de SPPAS. La table 2 rapporte le nombre de "tokens" de chacun des corpus après cette normalisation (colonne 2). Par

[1] http://www.sppas.org/. La version 1.9.4 a été utilisée pour cette étude.

[2] Enregistrements audio-vidéo et transcription orthographique sont disponibles sur Ortolang : https://hdl.handle.net/11403/sldr000875

[3] Enregistrements audio, transcription orthographique, et d'autres annotations sont disponibles sur Ortolang : https://hdl.handle.net/11403/sldr000720. Vidéo : https://hdl.handle.net/11403/sldr000027

token, nous entendons toute séquence transcrite, à savoir les mots ainsi que toutes les autres productions sonores. Les colonnes 3 à 5 indiquent le pourcentage que représentent respectivement les "euh", rires et bruits. On constate que chacun des corpus contient un pourcentage relativement élevé de "euh" : de 2,3% à 6% des tokens. Pour les corpus *Europe* (6%) et *CID* (4%), cela fait du "euh" le token le plus fréquent, largement devant "de" avec ses 4,22% dans *Europe* et le mot "est" avec 2,67% dans le *CID*. Dans le corpus *Typaloc*, comme dans *Europe*, c'est le mot "de" qui est le plus fréquent (3,03%), tandis que dans *Cheese*, c'est le mot "est" comme pour le *CID*, avec 3,06%. Avec 5,36% des occurrences, "tu" est le mot le plus fréquent du corpus *AixMapTask*. On en conclut que « euh » est très fréquent en parole spontanée, cependant sa fréquence dépend du style de parole : on l'observe moins dans les différents styles de conversations que dans un débat politique radio-diffusé.

Concernant les rires, le corpus *Europe* n'en contient qu'un seul, ce qui n'est pas surprenant compte-tenu du type de débat et du sujet abordé. Le rire est en revanche relativement fréquent dans le *CID*. Toutefois, c'est dans *Cheese* qu'on retrouve proportionnellement le plus grand nombre de rires ; il arrive en 3ème position des tokens les plus fréquents. Ainsi, c'est dans les deux conversations informelles qu'on retrouve le plus de rires. L'examen de ces résultats montre que plus la parole est relâchée, plus la fréquence des rires augmente. On trouve également un grand nombre de bruits, en particulier dans *AixMapTask,* corpus pour lequel ils représentent le 4ème token le plus fréquent. Nous nous abstenons toutefois de tirer une conclusion relative au style de parole car la présence de bruits tels que les inspirations/expirations dépend de la position/qualité du micro. Nous constatons cependant qu'il peut être fréquent dans les données.

Corpus	**Nombre de tokens**	**% de euh**	**% de rires**	**% de bruits**
Europe	7 566	6,014%	0,013%	0,264%
Typaloc	7 534	2,933%	0,186%	1,434%
AixMapTask	37 979	2,285%	0,635%	2,607%
CID	126 260	3,997%	1,221%	0,870%
Cheese	16 829	2,793%	2,246%	0,434%

TABLE 2 : Tokens et pourcentages que représentent les euh, rires et bruits

2.3 Contexte des euh, rires et bruits

La table 3 fait mention des contextes dans lesquels on retrouve les euh, rires et bruits. Effectivement, pour cette étude, dans laquelle nous nous intéressons plus particulièrement à la tâche de segmentation de la parole, il est utile de savoir dans quelle mesure le système automatique devra intervenir. La colonne 2 indique le pourcentage des euh, rires ou bruits qui sont entourés de silences, c'est-à-dire que l'événement constitue une IPU à lui seul. Dans ce cas, le système d'alignement forcé n'interviendra pas puisque l'événement a déjà été aligné par la segmentation en IPUs. Pour les colonnes 3 et 4, le système d'alignement forcé devra déterminer respectivement la frontière finale ou initiale de l'événement. Enfin, la dernière colonne indique les cas où l'événement se trouve entouré d'autres tokens (soit des mots, soit un autre événement), donc le système d'alignement forcé aura à déterminer ses frontières initiales et finales. On observe que plus d'un tiers des rires et un bruit sur cinq sont entourés de silences. Quant au "euh", on peut dire qu'il ne se retrouve quasiment jamais

isolé entre deux silences. Si l'on met en lien les tables 2 et 3, on en conclut que les euh, rires et bruits représentent un nombre important de tokens dans tous les corpus spontanés, mais dans des proportions différentes selon le style de parole.

	entouré de silences	**au début d'une IPU**	**à la fin d'une IPU**	**au sein d'une IPU**
euh	1,47%	11,80%	28,99%	57,75%
rire	34,65%	19,07%	29,09%	17,19%
bruit	21,19%	28,40%	11,84%	38,58%

TABLE 3 : Pourcentage des euh, rires et bruits en fonction de leur contexte gauche et droit

Pour terminer cette partie de l'étude des distributions des corpus, la table 4 indique le nombre et la proportion d'IPUs dans lesquelles on retrouve les euh, rires et bruits. Ces chiffres sont importants, en effet, puisque le système d'alignement forcé, qui opère séparément sur chacune des IPUs, utilise un algorithme d'optimisation global sur la séquence. A ce titre, il peut arriver qu'une erreur d'alignement affecte largement ses contextes droits et gauches. On voit ainsi que 20% à 36% des IPUs contiennent au moins un euh, un rire ou un bruit (dernière colonne).

Corpus	**# total IPUs**	**IPUS avec "euh"**	**IPUs avec rire**	**IPUs avec bruit**	**IPUs avec au moins un euh/rire/bruit**
Europe	875	35,88%	0,11%	2,29%	35,89%
Typaloc	522	28,25%	2,68%	14,94%	35,82%
AixMapTask	6126	12,16%	3,67%	13,52%	20,60%
CID	13631	27,32%	10,25%	7,52%	32,14%
Cheese	2675	14,62%	12,45%	2,73%	21,16%

TABLE 4 : Nombre d'IPUs dans lesquelles les euh, rires et bruits apparaissent

On constate également que ces événements se retrouvent très souvent dans les mêmes IPUs, puisque la dernière colonne est loin de représenter la somme des colonnes 2 à 4. Pour illustrer ce phénomène, nous avons extrait une IPU dans deux corpus :

- exemple de *Typaloc* : "donc *euh* des choses *euh* genre *euh* canard à l'orange des choses comme ça qui demandent *euh* une préparation un peu plus subtile une surveillance"
- exemple de *Cheese* : "tu vas avec ton père *euh* il repart avec mille chameaux à @"

3 Alignement forcé

3.1 Corpus de test et méthode d'évaluation

Un corpus de test a été manuellement phonétisé et aligné par un expert avec le logiciel Praat. Cette annotation a ensuite été vérifiée et éventuellement révisée par une seconde personne. Les fichiers de ce corpus ont été extrait aléatoirement du corpus CID. Il comprend 27 IPUs de 12 locuteurs différents, pour une durée totale de 141 secondes. En tout, 1833 labels devront être alignés par le système : 1791 phonèmes, 24 "euh", 5 rires, 4 bruits et 9 pauses courtes.

Les évaluations consistent à comparer la segmentation automatique à celle effectuée manuellement avec la mesure communément nommée "Unit Boundary Position Accuracy (UBPA)". Elle estime le pourcentage de frontières automatiques incluses dans une fenêtre d'une durée donnée autour des frontières manuelles correspondantes. C'est donc une mesure quantitative qui permet de situer globalement les performances d'un système, mais surtout, elle permet de comparer aisément et rapidement différents systèmes.

3.2 Résultats d'alignement avec la mesure UBPA

Pour réaliser cette étude, dans un premier temps, nous avons construit un modèle acoustique appris sur des données de parole lues. L'apprentissage a été réalisé à l'aide de la boîte à outils HTK (Young & Young, 1993) et de SPPAS, en suivant le tutoriel du site voxforge.org[4]. Un modèle HMM à 5 états du silence et de chacun des 31 phonèmes suivants ont ainsi été appris :

- voyelles : A/ E e 2 i O/ 9 u y
- voyelles nasalisées: a~ U~/ o~
- plosives : p t k b d g
- fricatives : f v s z S Z
- consonnes nasales : m n
- liquides : l R
- glides : H j w

pour lesquels A/ représente a ou A, O/ représente o ou O et U~/ représente e~ ou 9~, en SAMPA comme proposé par J.C. Wells[5]. Les euh, rires et bruits n'étant que peu présent voire absents des corpus lus, nous avons utilisé un modèle prototypique[6] pour chacun d'entre-eux. Le "euh" y est symbolisé par l'étiquette fp, le rire par lg et le bruit par gb. Par la suite, nous appellerons ce modèle "initial". Nous avons ensuite appris les modèles des euh, rires et bruits en suivant la même procédure d'apprentissage, que l'on a appliquée sur les corpus décrits dans la table 2. Ces derniers n'ont pas été utilisés pour l'apprentissage des modèles des phonèmes car le modèle appris à partir des données lues amène à des résultats significativement meilleurs (selon la mesure UPBA). Ces trois HMM ont alors été injectés dans le modèle initial, remplaçant ainsi les prototypes.

---

[4] A la différence du tutoriel, nous avons utilisé SPPAS pour normaliser, phonétiser et aligner les données pendant la phase d'apprentissage, en paramétrant SPPAS pour qu'il utilise *Julius* (Lee et al., 2001) plutôt que la commande *HVite* d'HTK lors de l'alignement. Nous avons utilisé la version 4.2.2 du "Open-Source Large Vocabulary CSR Engine Julius" : http://julius.osdn.jp

[5] http://www.phon.ucl.ac.uk/home/sampa/french.htm

[6] Modèle résultat de la commande *HCompV* d'HTK (version 3.4.1) : http://htk.eng.cam.ac.uk/

La table 5 indique les mesures UBPA avec une valeur de taille de fenêtre variant de 20ms à 80ms. Plusieurs modèles sont comparés : le modèle initial puis ce même modèle dans lequel on utilise soit le HMM prototype soit le HMM appris pour les euh, rires et bruits. Aucune modification n'est apportée aux autres HMM du modèle, et dans tous les cas, SPPAS est utilisé pour aligner (configuré pour appeler *Julius*). Ces résultats montrent que l'utilisation d'un modèle appris pour le bruit ne change pas les résultats. En revanche, l'introduction du HMM appris du rire améliore les performances, même dans ce corpus de test ne contient que 5 items. Enfin, l'introduction du HMM appris spécifiquement pour les euh augmente significativement la qualité de l'alignement.

	20ms	30ms	40ms	50ms	80ms
modèle initial (euh, rires et bruits prototypes)	84,91	92,16	94,32	95,45	97,02
avec HMM des bruits appris (euh et rires prototypes)	84,75	92,10	94,37	95,51	97,08
avec HMM des rires appris (euh et bruits prototypes)	85,13	92,48	94,75	95,94	97,67
avec HMM des euh appris (rires et bruits prototypes)	86,05	93,67	96,00	97,08	98,48
modèle final (euh, rires et bruits appris)	86,10	93,94	96,48	97,62	99,19

TABLE 5 : UBPA (%) du système d'alignement en utilisant différents modèles acoustiques

Pour compléter cette étude, nous ne pouvions pas ignorer le fait que les autres systèmes d'alignement forcé qui traitent la langue française utilisent la voyelle 2 pour aligner les "euh". Nous avons donc estimé les résultats en utilisant le modèle initial dans lequel le HMM du euh est remplacé par celui de la voyelle 2. Les mesures UBPA sont à 20ms : 85,67% ; à 30ms: 92,91% ; à 40ms : 95,19% ; à 50ms : 96,37% ; et à 80ms : 97,83%. Ainsi, l'utilisation de la voyelle 2 s'avère nettement plus judicieuse que l'utilisation du prototype, dans le cas où il ne serait pas possible de disposer d'un modèle spécifique. Nous avons néanmoins montré que l'apprentissage d'un modèle spécifique pour les "euh" amène à un meilleur résultat d'alignement significativement meilleur (avant dernière ligne du tableau) qu'en utilisant une voyelle acoustiquement proche.

3.3 Résultats qualitatifs des alignements

L'analyse qualitative des résultats présente un grand intérêt, notamment pour mieux comprendre et donc mieux appréhender les annotations obtenues. Nous nous sommes donc intéressés aux erreurs majeures, c'est-à-dire lorsque le système propose un désalignement qui dépasse 80ms, soit 15 cas dans notre corpus. Un point important à soulever concerne le fait que ces erreurs majeures se concentrent sur 5 IPUs seulement sur les 27 qui ont été alignées. La figure 1 illustre la séquence d'erreurs la plus importante du corpus, lors de l'alignement des tokens "na na na na na". Dans un segment de discours rapporté, le locuteur mentionne le fait que le discours continue mais ne présente pas d'intérêt pour le propos actuel. Bien qu'elle soit audible, il produit cette séquence de manière

assez hypo-articulée. Le système d'alignement échoue à déterminer les frontières entre les phonèmes n et A/ et 6 erreurs d'alignement sont ainsi observées sur cette suite de 12 phonèmes. La figure 2 illustre une autre cascade d'erreurs : le système ne détermine pas correctement le début du rire et lui assigne le phonème A/ du mot qui le précède, et cela affecte la suite des 4 phonèmes k-t-w-A/.

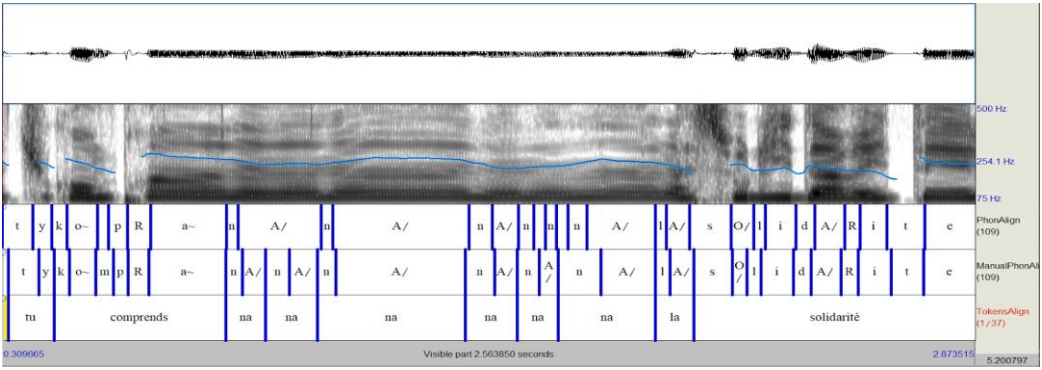

FIGURE 1 : Erreurs d'alignement sur la séquence de tokens "na na na na na na"

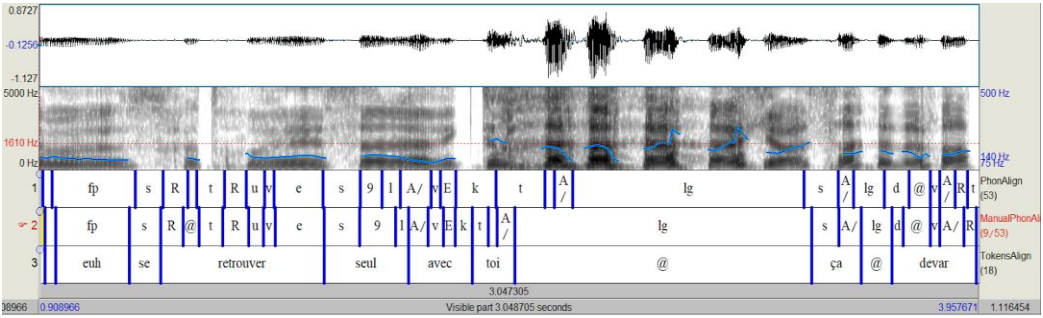

FIGURE 2 : Erreurs d'alignement sur la séquence de tokens "avec toi @"

4 Conclusion

Pour cette étude, nous avons rassemblé différents corpus de parole spontanée, en les sélectionnant de sorte qu'ils soient relativement homogènes dans leur forme (conditions d'enregistrement, transcription, etc). Ces corpus nous ont permis d'estimer les fréquences des tokens et de les comparer. Nous nous sommes intéressés en particulier aux *euh*, *rires* et *bruits*. Nous avons observé que le « euh » se retrouve fréquemment en parole spontanée, mais qu'il est nettement plus fréquent (6%) dans un débat politique que dans les conversations (2,3% à 4%). En revanche, plus la parole est relâchée, plus les rires sont présents ; ils peuvent en effet représenter jusqu'à 2,25% des tokens en conversation (sans consigne). On retrouve également une proportion non négligeable de bruits (produits par le locuteur) dans les différents corpus. Ces 3 items sont tellement fréquents en parole spontanée, qu'on les retrouve dans 20% à 36% des IPUs. En analysant leurs contextes gauche et droits, on constate qu'ils sont rarement isolés entre deux silences : seuls 1,47% des euh sont isolés. Ces résultats nous ont amené à évaluer l'impact de la prise en compte de ces 3 items dans la tâche d'alignement automatique : nous avons comparé l'utilisation d'un modèle acoustique qui inclut soit un HMM prototypique, soit un HMM appris. Nous en avons conclu qu'il n'était pas vraiment nécessaire d'apprendre un modèle de bruit. Des études plus poussées sur ce point nous permettraient

d'approfondir cet aspect. En revanche, la prise en compte des rires et surtout des euh, conduit à une nette amélioration des performances du système.

Les scripts et modules Python que nous avons développé pour l'apprentissage des modèles acoustiques ainsi que pour leur évaluation sont distribués sous licence GPL dans la version 1.9.4 de SPPAS et le modèle acoustique du français le sera dans la version 1.9.6.

Références

ADDA-DECKER M., GENDROT C., NGUYEN N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. Traitement Automatique des Langues, v. 49, n. 3, p. 13–46.

BERTRAND R., BLACHE P., ESPESSER R., FERRE G., MEUNIER C., PRIEGO-VALVERDE B., RAUZY S. (2008). Le CID — Corpus of Interactional Data — Annotation et Exploitation Multimodale de Parole Conversationnelle. Traitement Automatique des Langues, v. 49, n. 3.

BIGI B. (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician,* International Society of Phonetic Sciences, v. 111–112*,* p. 54–69.

GORISCH J., ASTÉSANO C., GURMAN BARD E., BIGI B., PRÉVOT L. (2014). Aix Map Task corpus: The French multimodal corpus of task-oriented dialogue. Proceedings of the 9th International conference on Language Resources and Evaluation, Reykjavik, Iceland, p. 2648–2652.

JOHNSON K. (2004). Massive reduction in conversational American English. In Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium, Tokyo, Japan, p. 29–54.

LEE A., KAWAHARA T., SHIKANO K. (2001). Julius -- an open source real-time large vocabulary recognition engine. Proceedings of the European Conference on Speech Communication and Technology, Aalborg, Denmark, p. 1691–1694.

MEUNIER C., FOUGERON C., FREDOUILLE C., BIGI B., CREVIER-BUCHMAN L. ET AL. (2016). The TYPALOC Corpus: A Collection of Various Dysarthric Speech Recordings in Read and Spontaneous Styles. Proceedings of the 10th Language Resources and Evaluation Conference, Portorož, Slovenia. p. 4658–4665.

MEUNIER C. (2013). Phoneme deletion and fusion in conversational speech. Proceedings of the Experimental Approaches to Perception and Production of Language Variation, Copenhague, Danemark.

PORTES C. (2004). Prosodie et économie du discours : spécificité phonétique, écologie discursive et portée pragmatique de l'intonation d'implication. Université de Provence - Aix-Marseille I (PhD).

PRIEGO-VALVERDE B., BIGI B. (2016). Smiling behavior in humorous and non humorous conversations: a preliminary cross-cultural comparison between American English and French. International Society for Humor Studies Conference, Dublin, Ireland.

YOUNG S.J., YOUNG SJ. (1993). The HTK hidden Markov model toolkit: Design and philosophy. University of Cambridge, Department of Engineering.