

Hésitations autonomes dans 8 langues : une étude acoustique et perceptive

Ioana Vasilescu¹, Maria Candea², Martine Adda-Decker³

¹ TELECOM PARIS, 46, rue Barrault, 75634 Paris cedex 13, ² Paris 3 - EA 1483 13 rue Santeuil, bur.431, 75005 Paris, ³ LIMSI-CNRS, Bât. 508, BP 133, 91403 Orsay cedex, France

Mél : vasilesc@tsi.enst.fr, candea@ext.jussieu.fr, madda@limsi.fr

ABSTRACT

This paper deals with the analysis of autonomous filled pauses in 8 languages. The general question addressed concerns whether autonomous filled pauses carry language-specific information. They occur frequently in spontaneous speech and represent an interesting topic for improving language-specific models for automatic language identification. Most of the current studies focus on few languages such as English and French. Our study aims to describe the acoustic peculiarities of autonomous filled pauses in French, English, Spanish, Italian, Portuguese, German, Mandarin Chinese and Middle-Oriental Arabic. Thus we propose an acoustic typology based on the vocalic characteristics of autonomous filled pauses. We also compare the vocalic segments of filled pauses with intra-lexical vocalic segments possessing similar acoustic peculiarities. In this purpose, a preliminary study on French is described here. Finally, a perceptual experiment concerning French/L2 discrimination based on filled pauses stimuli is discussed.

1. INTRODUCTION

Dans les corpus de parole la proportion des phénomènes dits d'hésitation ou de formulation peut varier de manière très significative suivant le style de parole. Dans la parole lue et préparée (journaux télévisés ou radiodiffusés) leur taux reste en général inférieur à 0,5% des mots. En revanche dans la parole conversationnelle la proportion de tels phénomènes peut atteindre facilement des taux de 5-10%. Dans les études portant sur la parole non lue, la modélisation des phénomènes d'hésitation devient indispensable. C'est d'ailleurs devenue chose courante dans les systèmes de transcription automatique de la parole depuis que les styles de parole traités se sont diversifiés en s'éloignant des simples corpus de textes lus [9].

Les phénomènes dits d'hésitation sont multiples variant de simples allongements de syllabe, de pauses silencieuses ou d'insertions de voyelles autonomes, aux reprises et auto-corrections plus complexes. De nombreux auteurs s'intéressent aux phénomènes d'hésitation soit d'un point de vue purement linguistique [2, 3, 7] soit en adoptant une perspective plus orientée vers le traitement automatique [8, 13]. La plupart des études à base de grands corpus ont été faites sur du français ou de l'anglais et il a été constaté dans ces deux langues que la voyelle autonome

d'hésitation, pouvant être insérée a priori dans toute position, est une voyelle centrale fortement allongée. Nous n'avons pas trouvé d'étude approfondie sur les différences de timbre entre les voyelles d'hésitation de ces deux langues et nous n'avons pas non plus trouvé d'étude détaillée sur la possibilité de trouver, à travers les langues du monde, des voyelles plus antérieures ou plus postérieures pouvant jouer ce même rôle.

L'étude des phénomènes d'hésitation dans un cadre multilingue est une perspective relativement nouvelle. Ce travail exploratoire s'inscrit dans le projet MIDL (Modélisations pour l'identification des langues) faisant partie du programme interdisciplinaire du CNRS STIC-SHS et réunissant les laboratoires LIMSI-CNRS, LTCI-ENST, CTA/DGA, LPP-ILPGA Paris 3 et EA1483 Paris 3. Le but du projet MIDL est de mettre ensemble des compétences pluridisciplinaires linguistiques et informatiques afin d'augmenter nos connaissances sur l'identification des langues et éventuellement contribuer à améliorer la modélisation pour l'identification automatique des langues.

Nous nous intéressons ici aux voyelles d'hésitation autonomes (appelées parfois pauses remplies) dans huit langues et nous essayons principalement de répondre à la question si ces hésitations autonomes sont porteuses d'information spécifique à la langue. D'autres questions sont également considérées à savoir : le timbre des hésitations autonomes est-il toujours central ? Correspond-il à une voyelle du système vocalique de la langue en question ? Si oui, y a-t-il des paramètres, autres que la durée, qui permettraient de différencier la voyelle d'hésitation de son homologue du système vocalique (valeur de F0, contour de F0, qualité de voix, énergie...)?

Nous tâcherons d'apporter quelques réponses à ces questions par des tests perceptifs et des analyses acoustiques. Deux corpus seront présentés et étudiés : tout d'abord un corpus multilingue et ensuite un corpus de français qui nous servira à affiner nos observations notamment pour ce qui est du lien entre la/les voyelle(s) d'hésitation et la/les voyelle(s) intra-lexicales de timbre similaire de la langue. Suivant les réponses aux questions posées, la modélisation plus fine de ces phénomènes pourrait être intéressante pour l'identification automatique des langues.

2. ANALYSE ACOUSTIQUE ET PERCEPTIVE

2.1. Présentation du corpus multilingue

Notre étude part de corpus journalistiques avec de la parole relativement bien préparée : on y trouve peu de phénomènes de disfluences et ils se limitent essentiellement aux allongements vocaliques de mots-outils et aux hésitations vocaliques autonomes. S'agissant d'une étude exploratoire, un corpus bien contrôlé a été préféré. Dans un deuxième temps l'étude sera étendue à des corpus de style d'élocution plus libre qui permettront d'aborder la question de la variation éventuelle des hésitations autonomes en fonction du style de parole. Pour les huit langues étudiées ici (Français, Espagnol, Italien, Portugais, Anglais, Allemand, Arabe moyen-oriental et Chinois mandarin) nous avons extrait de manière semi-automatique une centaine d'hésitations autonomes par langue. Toutes ces voyelles extraites ont été vérifiées à la main afin qu'il ne s'agisse pas d'allongements vocaliques de mots-outils. Plus précisément, nous avons employé au minimum 32 (en anglais) et au maximum 209 (en français) hésitations autonomes par langue (donc en moyenne : 80 hésitations autonomes/langue). Les hésitations sont prononcées par des locuteurs hommes et femmes, de l'ordre d'une dizaine en moyenne par langue. Malgré la taille relativement petite encore de notre corpus d'hésitations multilingue, il nous permet d'avoir un premier aperçu des caractéristiques acoustiques du phénomène. Nous nous proposons de faire appel dans nos futures études à beaucoup plus de données, équilibrées en termes d'échantillons et de nombre de locuteurs distincts par langue.

2.2. Définition de l'hésitation autonome

L'objet de cette étude est la voyelle d'hésitation autonome : il s'agit d'un segment vocalique allongé inséré entre deux pauses ou bien à la suite d'un mot mais ne constituant pas un allongement de celui-ci. Ce type de segment, appelé parfois pause remplie, présente généralement une durée allant bien au-delà de la durée typique d'une voyelle habituelle. Nous avons préféré restreindre cette étude à l'observation d'hésitations autonomes extra-lexicales précédées par une pause silencieuse, pour éviter tout problème d'ambiguïté de segmentation. Un critère de durée a été adopté pour la sélection (durées vocaliques supérieures à 200 ms).

A partir des hésitations autonomes ainsi isolées nous cherchons à établir une typologie acoustique basée sur le timbre des voyelles de ces hésitations. Cette voyelle est parfois diphtonguée à la fin ou présente une coda nasale. Nous appelons alors *voyelle support*, la voyelle la plus longue et la plus stable de chaque occurrence. La voyelle support constitue l'élément principal d'une hésitation. Un test de perception a ensuite été élaboré afin de vérifier si l'oreille humaine est capable de

distinguer des différences acoustiques observées préalablement.

2.3. Etude acoustique des voyelles support

Une étude préliminaire du corpus multilingue [4] a permis de répondre, au moins partiellement, à la question de départ, à savoir si l'hésitation autonome porte une information spécifique à la langue. Ainsi, l'analyse acoustique de la distribution des voyelles support dans un espace bidimensionnel F1/F2 a montré que la dispersion inter-langue est plus importante que celle intra-langue. Suivant les langues analysées on peut trouver des voyelles support différentes qui sont loin d'être toujours des voyelles centrales mi-fermées. Les voyelles utilisées se placent sur deux axes, central et antérieur, allant de [æ] à [e] en passant par le [œ]. Nous n'avons pas identifié de réalisations postérieures parmi les voyelles analysées jusqu'à ce jour. Suivant les langues la dispersion peut être plus ou moins grande : alors que l'hésitation autonome du français s'organise bien autour d'une voyelle centrale arrondie (qui correspond bien à sa transcription lexicalisée *eu*h), l'italien admet des hésitations autour du timbre [e] mais également autour de la voyelle centrale [œ], qui n'est pas présente dans son système vocalique. Toutes les langues étudiées ici admettent cependant comme voyelle support au moins une voyelle présente dans leur système vocalique.

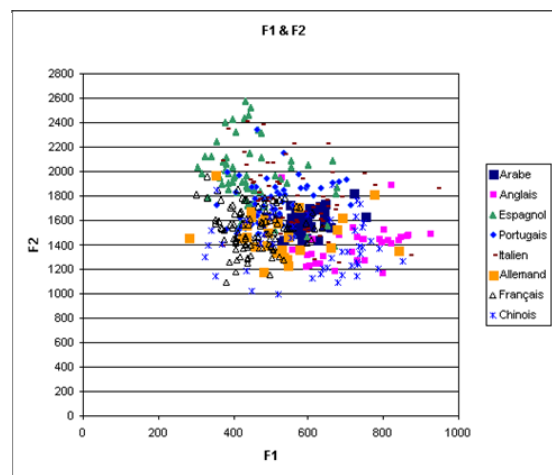


FIG. 1 – Dispersion des voyelles support des hésitations en 8 langues dans le plan F1-F2.

2.4. Etude perceptive

Afin de valider ces informations acoustiques un test perceptif a été mené. Dans notre étude préliminaire [4], les sujets étaient tous francophones, sans être nécessairement de langue maternelle française, et les stimuli ont été sélectionnés surtout par rapport au statut d'autonomie de l'hésitation. Ici des stimuli pour un nouveau test français/L2 ont été sélectionnés de façon beaucoup plus contrôlée. Des variables telles que la

qualité de voix et le timbre de la voyelle ont été considérées lors de la sélection. Nous avons utilisé exclusivement des hésitations autonomes non-enchaînées, dans toutes les langues. Les stimuli ne présentent pas des qualités de voix extrêmes pouvant servir d'indice de reconnaissance et couvrent les timbres vocaliques révélés par l'analyse acoustique comme étant typiques de chaque langue du corpus. Ces stimuli ont ensuite été soumis à des auditeurs de langue maternelle française. 20 sujets français ont participé.

Les résultats sont illustrés par la figure 2 ci-dessous. Ils montrent qu'en moyenne les hésitations autonomes des sept langues du corpus ont été discriminées des hésitations du français avec un taux de réussite moyen de 75%. Les taux de reconnaissance sont significativement supérieurs au hasard pour toutes les paires de langues. Les scores les plus bas sont obtenus pour les trois paires de langues Français/Arabe, Français/Portugais et Français/Allemand (<65%), tandis que les quatre paires de langues Français/Italien, Français/Espagnol, Français/Anglais et Français/Chinois obtiennent des scores supérieurs à 80%.

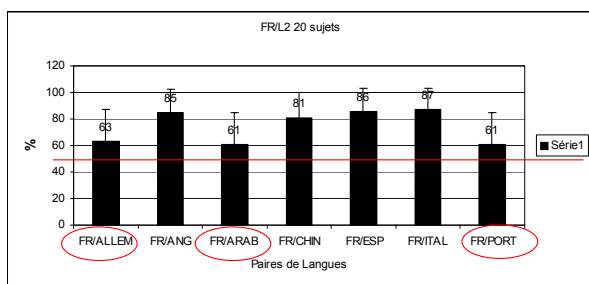


FIG. 2 – Pourcentages de réussite en discrimination Français/L2.

A partir de l'expérience préliminaire [4], nous avons pu formuler l'hypothèse que la langue maternelle ne joue pas un rôle prépondérant dans la discrimination, même si la connaissance *a priori* des hésitations autonomes de la langue maternelle a pu aider les auditeurs dans leurs décisions. Cette hypothèse est soutenue par le fait que les langues ayant des hésitations autonomes de timbre similaire au français (arabe, allemand, portugais) entraînent des taux de confusion élevés pouvant aller jusqu'à 39% en moyenne alors que les langues dont la voyelle support est de timbre différent à celle du français (par exemple, l'espagnol) sont discriminées à plus de 80%. De plus, des tests statistiques (t-test appariés, $p < 0,0001$) montrent que les auditeurs ont significativement mieux reconnu l'anglais, l'espagnol, l'italien et le chinois que l'arabe, le portugais et l'allemand par rapport au français. Nous avons également évalué les rôles des facteurs « langue maternelle » et « L2 » dans la discrimination. Il a été constaté un effet significatif du facteur « L2 » dans les scores de discrimination (ANOVA, $p < 0,0001$). « L2 »

est globalement et statistiquement mieux reconnu que le français (ANOVA, $p = 0,0198$) notamment pour les paires Français/Arabe, Français/Chinois et Français/Allemand. Nous posons l'hypothèse que les auditeurs ont cherché dans les hésitations ce qui était différent du prototype *a priori* d'hésitation française.

A l'issue de ce test et afin de corroborer les résultats obtenus, un test L1/L2 où L1 est une langue autre que le français (possédant éventuellement une hésitation dont la voyelle support a un timbre vocalique de type central/proche du français) nous semble nécessaire afin de circonscrire le rôle de la connaissance *a priori* des hésitations de la langue maternelle dans la discrimination correcte.

3. COMPARAISON VOYELLE D'HESITATION VS. VOYELLES DU SYSTEME VOCALIQUE

Nous nous sommes proposé d'étudier s'il existe un rapport entre voyelles d'hésitations et voyelles du système vocalique. Nous avons vu (cf. 2.3) que l'italien admet une voyelle d'hésitation autour de la voyelle centrale [œ], qui n'est pas présente dans son système vocalique. Toutes les huit langues étudiées admettent cependant comme voyelle support au moins une voyelle présente dans leur système vocalique. Nous avons mené une étude plus approfondie sur ce sujet pour l'instant uniquement sur le français en prenant en considération toutes les voyelles intra-lexicales de timbre similaire à la voyelle support de l'hésitation.

3.1. Présentation du corpus français étendu

En français les voyelles support d'hésitation correspondent à une voyelle centrale plutôt arrondie. Dans le système vocalique français nous trouvons dans cette zone le « eu-fermé », le « eu-ouvert » et le « schwa ».

Pour cette étude le corpus français a donc été étendu, incorporant davantage d'hésitations et les voyelles intra-lexicales de timbre similaire. Le corpus français étendu comprend ainsi les données utilisées dans [4], mais aussi d'autres occurrences d'hésitations autonomes, isolées par des pauses ou enchaînées au mot précédent, prononcées par différents locuteurs, hommes et femmes, des occurrences de voyelles de timbre similaire « schwa »¹ [ə] et « eu-ouvert » [œ] confondus ici, et des « eu fermé » [ø], dans différents

¹ Dans le cadre du traitement automatique nous rassemblons sous le nom de voyelle « schwa » à la fois des occurrences de « schwa » tel qu'il est défini dans les dictionnaires, à savoir une voyelle qui peut être prononcée ou non selon le contexte et le registre, et les « eu-ouvert » [œ], ces derniers étant cependant beaucoup moins nombreux. Lorsque les schwas sont réalisés, leur timbre est très proche, voire indifférenciable, de celui du [œ].

contextes. Les segments vocaliques ont été extraits à l'aide d'un alignement automatique effectué par le système de reconnaissance du LIMSI à partir d'enregistrements journalistiques diversifiés (radio et télévision) [9]. Les segments retenus présentent tous une durée minimale supérieure à 40 ms pour les voyelles intra-lexicales. De manière générale, les locuteurs masculins sont beaucoup plus nombreux dans les corpus journalistiques que les locuteurs féminins (en moyenne 70% locuteurs masculins vs. 30% locuteurs féminins). Cette observation se retrouve dans la proportion des hésitations et des « eu fermé » [ø] dans la table 1 ci-dessous. La voyelle [ø] est en gros dix fois moins fréquente en français que sa contrepartie ouverte [œ-ø]. En ce qui concerne cette voyelle-composite [œ-ø], nous n'avons retenu qu'un millier d'occurrences, même si la proportion *a priori* au niveau du corpus reste la même (70% vs. 30% avec 10k d'occurrences pour les hommes et 3k pour les femmes).

TAB. 1 – Nombre d'occurrences des voyelles d'hésitation et des voyelles de timbre similaire dans le corpus français étendu.

	hommes	femmes
hésitations	1484	378
eu-ouvert [œ - ø]	964	754
eu-fermé [ø]	923	291

3.2. Analyse acoustique

Nous nous sommes posé essentiellement la question de savoir si on peut distinguer, par des mesures acoustiques, les voyelles support d'hésitation des voyelles intra-lexicales.

Plusieurs paramètres acoustiques sont pris en compte :

- la durée des segments,
- le timbre de la voyelle (distribution des segments vocaliques dans l'espace bidimensionnel F1/F2),
- la hauteur et le contour de la fréquence fondamentale F0,
- la qualité de voix,

L'analyse de nos données confirme les observations d'études antérieures comme [2] ou [3] quant à la **durée** : la distribution de la durée des hésitations se distingue de celle des autres voyelles surtout par une proportion très importante d'occurrences dépassant 200ms dans le cas des hésitations (voir figures 3a, 3b et 3c ci-après). Même si l'on retrouve des hésitations dont la durée se situe en dessous du seuil de 200ms que nous avons sélectionné, leur distribution de ce point de vue occupe préférentiellement une étendue de durée allant de 200 à 650ms comme la figure 3a le montre. En revanche les voyelles intra-lexicales atteignent rarement 200ms.

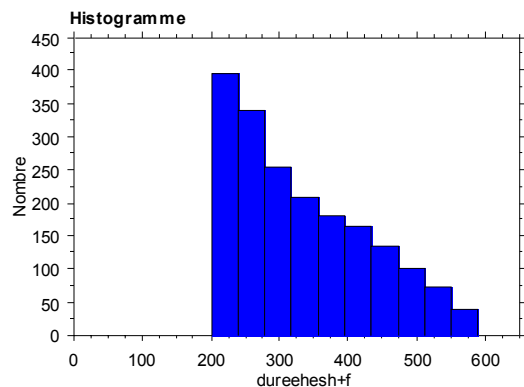


FIG. 3a – Distribution des hésitations en fonction de la durée (hommes+femmes).

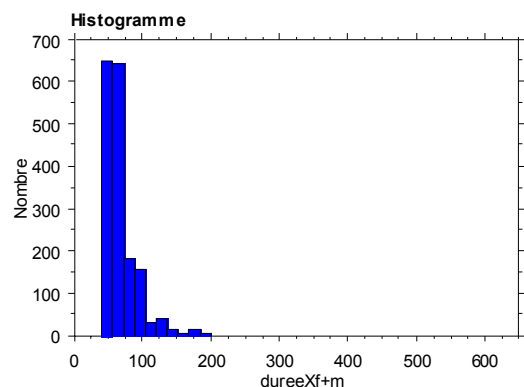


FIG. 3b – Distribution des schwa en fonction de la durée (hommes+femmes).

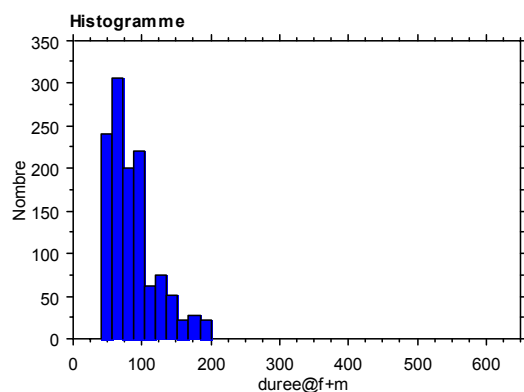


FIG. 3c – Distribution des eu-fermés en fonction de la durée (hommes+femmes).

Nous nous sommes intéressées par la suite au **timbre** des différents segments vocaliques analysés. A cet effet

nous avons fait appel au logiciel PRAAT². Les mesures ont été prises pour chacun des deux premiers formants au milieu (partie stable) de chaque segment vocalique. L'étude du timbre de ces voyelles à travers les mesures des deux premiers formants F1, F2 montre que F2 est sensiblement le même pour les différents cas de figure (voir table 2) Nous avons rajouté, dans table 2, des valeurs de formants canoniques, telles qu'elles sont utilisées dans [6] à partir de valeurs considérées comme références pour la langue française et fournies par G. Fant et dans Calliope[1].

Sur l'axe F1 (ouvert/fermé) un écart assez important entre les voyelles d'hésitations et celles du système vocalique (voir table 2) est mesuré. Pour les hommes, ces différences sur l'axe ouvert/fermé concernent les valeurs moyennes des trois catégories de segments vocaliques mesurés. Ainsi, les hésitations sont plus ouvertes que les « schwa »/« eu-ouvert » et que les « eu-fermé ». A leur tour, les « schwa »/« eu-ouvert » sont plus ouvertes que les « eu-fermé ».

Les mesures du premier formant F1 pour les voyelles d'hésitation sont proches des valeurs de référence du eu-ouvert [œ], pour les hommes et les femmes, telles qu'elles sont préconisées dans la littérature [1,12]. En revanche les mesures du F1 pour les voyelles du système vocalique proprement dit s'écartent de manière significative des valeurs de référence. Cet écart est à mettre en relation avec la durée (voir [6] pour plus de détails), significativement plus courte ici (autour de 60-80 ms) que pour les hésitations (>200ms).

TAB. 2 – Les valeurs moyennes des F1 et F2 pour les voyelles support des hésitations autonomes, les schwa et les eu-fermé en français (en Hz).

	F1	F2
hésitations H / F	470 / 523	1464 / 1659
[ə - œ] H / F	404 / 523	1421 / 1659
réf. Calliope-Fant	500 / 550	1450 / 1650
[ø] H / F	382 / 430	1465 / 1666
réf. Calliope-Fant	400 / 450	1450 / 1650

Nous nous sommes demandées si l'étude de l'axe étiré/arrondi (F3) permettrait une meilleure distinction des segments analysés ici. Nous nous appuyons sur les résultats de [6] qui trouvent des valeurs moyennes autour de 2500Hz pour « eu-fermé » et pour « schwa/eu-ouvert ». Il semblerait donc que F3, détecté automatiquement, ne permette pas de discriminer entre des segments [ø] et [œ - ə] tels qu'ils sont produits en parole spontanée.

² Les mesures des formants ainsi que celles de la F0 ont été faites avec le logiciel PRAAT, www.praat.org développé par P. Boersma et D. Weenink.

Concernant la **fréquence fondamentale** F0 nous montrons dans la table 3 la valeur moyenne et l'écart type pour chaque type de voyelle. Les valeurs moyennes de F0 sont assez similaires entre hésitations autonomes et voyelles. Les mesures de la table 3 montrent un F0 moyen plutôt élevé pour les hommes et plutôt faible pour les femmes. Cette observation ne permet donc pas de conclure que les voyelles d'hésitation sont globalement plus graves que les voyelles similaires du système vocalique. Cette question est à affiner, en particulier en étudiant les segments rejetés par le détecteur de f0.

TAB. 3 – F0 (en Hz) pour les voyelles support des hésitations autonomes, des schwa et des eu-fermés en français).

	F0Moy	E-Type
hésitations hommes	155	97,3
femmes	179	68,5
schwa hommes	114	48,6
femmes	192	38,3
eu-fermé hommes	144	45,4
femmes	219	56

Les résultats préliminaires montrent cependant que les hésitations présentent une étendue de la distribution plus large avec plus de valeurs « extrêmes », et cela autant chez les hommes que chez les femmes. Les valeurs élevées semblent plutôt correspondre à des erreurs de détection : une écoute des hésitations n'a révélé que très peu de segments à pitch aigu (mise à part une voix d'un jeune garçon, comptabilisé dans les voix d'hommes). En revanche, de nombreux segments présentent une hauteur perçue plutôt grave et une **qualité de voix** irrégulière (*vocal fry*). Ce type de segment peut poser des problèmes aux algorithmes d'extraction de pitch.

Nous avons donc par la suite évalué la proportion de segments pour laquelle l'extraction de F0 a échoué. La non détection de la F0 concerne plus souvent les voyelles d'hésitation, ce qui renforce l'hypothèse que les hésitations sont plus irrégulières que le reste de la parole. Ainsi, la non détection de la F0 avec PRAAT concerne 3% des productions des hommes et 0,5% des productions des femmes pour ce qui est des voyelles intra-lexicales (« schwa »/« eu-ouvert » et « eu-fermé »). Pour ce qui est des hésitations, les erreurs concernent 11,5% des productions des hommes et 8,2% des productions des femmes. Ces observations vont dans le sens d'une qualité de voix plus fluctuante (voix de type *vocal fry* et/ou *creaky*, éventuellement voix de sifflet, etc.) pour les hésitations par rapport à la parole continue non-hésitante, ce qui va dans le sens des observations de [8] et [10] pour l'anglais américain ; les voyelles intra-lexicales sont plutôt modales. Dans une phase ultérieure de nos travaux nous nous proposons d'approfondir l'étude de la qualité de voix des hésitations par rapport au reste de la parole. Si nos

hypothèses se confirment, elles pourraient être mises en relation avec un manque de contrôle (en particulier du souffle) ou d'irrégularité de la phonation en situation d'hésitation, liés à l'allongement significatif de la durée totale et à la baisse de F0. Nous n'avons pas encore analysé la forme des contours de F0, qui peuvent également porter une information permettant de distinguer les hésitations des voyelles intra-lexicales.

Cet axe d'étude, explorant le lien entre les voyelles d'hésitation et les voyelles du système vocalique est à poursuivre pour d'autres langues.

4. CONCLUSIONS

Nos conclusions acoustiques partielles ainsi que les tests de perception menés sur des sujets français plaident d'ores et déjà en faveur de l'hypothèse de l'existence d'informations spécifiques à la langue dans les voyelles d'hésitation autonomes et en faveur d'une meilleure modélisation de ces phénomènes pour une exploitation automatique dans l'identification des langues.

La discussion de nos données sur le français ouvre des pistes de réflexion intéressantes sur la relation qui existe entre les voyelles support des hésitations autonomes et les voyelles répertoriées dans le système vocalique et permet d'envisager les éventuelles caractéristiques universelles des hésitations à partir des huit langues étudiées.

Il reste néanmoins de nombreuses questions à affiner, comme par exemple les hypothèses qui permettraient de faire la distinction entre une hésitation autonome enchaînée et des mots outils allongés. Cette question se pose pour toutes les autres langues de la même manière que pour le français. Dans certaines langues, la voyelle d'hésitation autonome peut également constituer un mot qui existe dans la langue (une conjonction par exemple en italien, ou un déictique en japonais [11]) et il serait intéressant de savoir s'il existe des différences acoustiques pertinentes entre ces deux cas de figure.

REMERCIEMENTS

Nous remercions ici Cédric Gendrot pour son aide à l'élaboration de scripts sous PRAAT.

RÉFÉRENCES

- [1] Calliope. *La Parole et son Traitement Automatique*. Paris : Masson, 1989.
- [2] M. Candea, *Contribution à l'étude des pauses silencieuses et phénomènes dits "d'hésitation" en français oral spontané*. PhD dissertation, Université Paris 3, 2000.
- [3] H. H. Clark, Managing problems in speaking. *Speech Communication*, 15, 243-250, 1994.
- [4] J. Clerc-Renaud, I., Vasilescu, M., Candea & M. Adda-Decker. Etude acoustique et perceptive des

hésitations autonomes multilingues, *XXV^{es} JEP*, Fès Maroc, 2004.

- [5] D. Duez. "Acoustico-phonetic characteristics of filled pauses in spontaneous French speech: preliminary results", In *DISS'01*, 41-44, 2001.
- [6] C. Gendrot, M. Adda-Decker. Analyses formantiques de corpus d'émissions journalistiques en français et allemand, *MIDL2004*, Paris, 29-30 nov. 2004.
- [7] P. Léon. E caduc : facteurs distributionnels et prosodiques dans deux types de discours, *11th ICPHs*, 1987.
- [8] E. Shriberg. To 'errrr' is human: ecology and acoustics of speech disfluencies, *Journal of the International Phonetic Association*, 31/1, 2001.
- [9] J.L. Gauvain, L. Lamel, G. Adda., The Limsi Broadcast News Transcription System, *Speech Communication*, 37(1-2):89-108, 2002.
- [10] E. Shriberg, Phonetic consequences of speech disfluency, in *ICPhS*, San Francisco, 1999.
- [11] M. Watanabe, The constituent complexity and types of fillers in Japanese, *15th ICPHs*, Barcelone, 2003.
- [12] G. Fant (1973) *Speech sounds and features*, MIT Press, Cambridge, USA.
- [13] M. Adda-Decker, B. Habert, C. Barras, G. Adda, P. Boula de Mareuil, P. Paroubek, Une étude des disfluences pour la transcription automatique de la parole spontanée et l'amélioration des modèles de langage, *XXV^{es} JEP*, Fès Maroc, 2004.