

---

# Caractéristiques acoustiques et prosodiques des hésitations vocaliques dans trois langues

Ioana Vasilescu — Martine Adda-Decker — Rena Nemoto

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur,  
LIMSI-CNRS, BP.133, F91403 Orsay cedex

{ioana, madda, nemoto}@limsi.fr

---

*RÉSUMÉ. Cette contribution porte sur l'analyse des paramètres acoustiques et prosodiques caractérisant les hésitations vocaliques dans trois langues. Le but principal est de répondre à la question suivante : les hésitations vocaliques partagent-elles des traits spécifiques à la langue ? Nous nous interrogeons également sur le lien entre voyelles d'hésitation et système (vocalique) de la langue. Enfin, cette recherche est menée sur de grands corpus oraux permettant une généralisation des observations et la mise en évidence potentielle d'universaux linguistiques. Les voyelles d'hésitation et intralexicales sont extraites automatiquement de corpus consistant en quelques dizaines d'heures d'émissions radiodiffusées et de débats parlementaires en anglais américain, français et espagnol. Les paramètres mesurés et comparés sont le timbre à travers les formants F1 et F2, la durée et la fréquence fondamentale. Les résultats montrent que les hésitations vocaliques partagent certains traits (i.e., durée, hauteur), mais déploient aussi des spécificités liées à la langue, notamment en terme de qualité vocalique. Nos données montrent que les voyelles d'hésitation présentent des timbres différents à travers les langues. Ces timbres sont proches de timbres intralexicaux sans nécessairement se confondre avec eux.*

*ABSTRACT. This contribution provides a study on the acoustic and prosodic characteristics of vocalic hesitations in three languages. The main goal is to determine whether vocalic hesitations can be considered as bearing language-specific information. An additional point of interest concerns the link between vocalic hesitations and the vowels in the phonemic inventory of a language. Finally, a complementary aim of the present work is to use large corpora to investigate whether some language universals can be found. Hesitations have been automatically extracted from large corpora of journalistic broadcast speech and parliamentary debates in American English, French and Spanish. Duration, fundamental frequency and formant values were measured and compared. Results confirm that vocalic hesitations share (potentially universal) properties across languages, e.g., longer durations and lower fundamental frequency than are observed for intra-lexical vowels. The results on vocalic timbre show that while the measures on hesitations are close to existing vowels of the language, they do not necessarily collapse with them.*

*MOTS-CLÉS : hésitations vocaliques, dysfluences, systèmes vocaliques, formants, durée, fréquence fondamentale, universaux.*

*KEYWORDS : vocalic hesitations, speech disfluency, vocalic systems, formants, duration, fundamental frequency, language universals.*

---

## 1. Introduction

Le terme « hésitation » désigne dans la linguistique de corpus une variété de phénomènes vocaux, allant de vocalisations simples telles que les pauses remplies ou hésitations vocaliques, les allongements vocaliques, jusqu'à des marques de formulation plus complexes comme la répétition. Parmi ces phénomènes, les hésitations vocaliques, typiquement « euh » en français, ont bénéficié d'une attention toute particulière. Durant le dernier demi-siècle, elles ont fait l'objet d'analyses relevant de disciplines différentes telles que la psycholinguistique, la linguistique clinique, les sciences cognitives, la sociolinguistique, etc. (O'Connell *et al.*, 2004).

L'intérêt pour le statut et la fonction des hésitations vocaliques dans la langue est ancien. Dans son célèbre ouvrage « *Language* », L. Bloomfield est le premier à noter leur présence dans la langue parlée (Bloomfield, 1933). Cependant, ce n'est qu'en 1959 que les hésitations vocaliques deviennent un véritable objet d'étude, grâce au travail fondateur des psycholinguistes H. Maclay et C.-E. Osgood qui établissent une distinction entre deux types de parole, *fluente* et *dysfluente*, cette dernière se caractérisant par la présence de pauses, qu'elles soient silencieuses ou remplies de vocalisations. Dès lors, les pauses remplies appelées aussi « hésitations vocaliques » sont associées aux marques de « dysfluences » - des événements non lexicaux généralement irréguliers - et à la parole mal formée, en opposition avec une version « fluente » et « bien formée » du discours oral (Maclay *et al.*, 1959). La principale fonction pragmatique (communicative) des hésitations vocaliques dans ce type de parole est de signaler le désir du locuteur de garder son tour de parole. Par la suite, les hésitations vocaliques en tant que type de pause ont été associées à tout un éventail de rôles de planification en lien avec l'effort du locuteur à construire son message verbal dans le temps (Clark, 2002). Des expérimentations différentes dans la parole normale et pathologique ont montré que cet effort peut se traduire à différents niveaux linguistiques : lexical (Goldman-Eisler, 1968), lexical et/ou syntaxique (Grojean *et al.*, 1975 ; Boomer, 1965 ; Duez, 1991 ; etc.) ou plus général, au niveau de la « préparation » du message même (Clark *et al.*, 2002). Il peut être associé à la gestion des tours de parole (Maclay *et al.*, 1959) ou à une marque pragmatique du sentiment de confiance du locuteur dans ce qu'il dit (feeling of knowing) ou de l'auditeur à propos de la confiance du locuteur en sa réponse (feeling of another's knowing) (Brennan *et al.*, 1995). D'autres auteurs, comme G.-V. Mahl ont mis en relation hésitations vocaliques et émotions (Mahl, 1954) ou associé production des hésitations vocaliques et troubles du langage et, plus précisément, l'aphasie (Butterworth, 1979). Enfin, les études en amont menées en psychologie cognitive ont mis en relation production de pauses remplies ou silencieuses et complexité du traitement cognitif de l'information. Ainsi, le locuteur a un but communicatif à atteindre et les hésitations témoignent de son « effort de mise en mots de la pensée » (Chafe, 1980).

Pour ce qui est du statut des hésitations vocaliques dans la langue, la définition des psycholinguistes H. Maclay et C.-E. Osgood, selon laquelle les hésitations vocaliques sont des vocalisations sans contenu lexical, s'est imposée plusieurs décennies durant. Des études plus récentes sur l'anglais américain ont cependant fourni des éléments pour un réajustement de cette perspective en faveur d'une hypothèse qui rapproche les hésitations vocaliques du niveau lexical d'une langue (Clark *et al.*, 2002). Plus précisément, H. Clark et J. Fox Tree assimilent les hésitations vocaliques aux interjections. Elles sont également représentées comme une entrée lexicale entière dans les listes de mots utilisés pour la transcription automatique de la parole. Si l'appartenance des hésitations vocaliques à la classe des interjections est contestée par d'autres chercheurs tels que D.-C. O'Connell et S. Kowall, leur caractère (presque) lexical, de « vocalisations systématiques » est pleinement accepté (O'Connell *et al.*, 2005).

Nous nous intéressons ici aux hésitations vocaliques autonomes dans différentes langues. Nous appelons ici « hésitation vocalique » l'insertion dans le flux verbal d'une voyelle fortement allongée, telle que « euh » en français. Nous opérons une distinction entre voyelles d'hésitation et voyelles intralexicales allongées (par exemple, « la\_euh » vs. « laaaaa »), ces dernières étant considérées comme faisant partie intégrante du mot. Nous notons également des réalisations différentes des hésitations vocaliques : elles peuvent être attachées à un mot (« donc\_euh ») ou autonomes, séparées par des pauses silencieuses du contexte environnant (« donc//euh ») ; elles peuvent également être simplement vocaliques ou présenter une structure plus complexe, par exemple, une coda consonantique nasale (comme l'hésitation « um » en anglais américain). Les hésitations vocaliques sont particulièrement courantes dans la parole, pouvant atteindre 5 % des données suivant les styles de parole (Vasilescu *et al.*, 2005). Enfin, les transcriptions orthographiques des hésitations vocaliques varient à travers les langues, par exemple, « euh » en français, « eh » en espagnol, « uh/um » en anglais américain, ou bien « er » en anglais britannique, voire au sein d'une langue (*cf.* O'Connell *et al.*, 2004, pour les différentes transcriptions en anglais américain) suggérant des différences dans leur perception par les auditeurs natifs.

Si le rôle des hésitations vocaliques au sein de la communication verbale fait maintenant consensus grâce aux nombreuses études qui ont considéré le phénomène du point de vue de la production, les études consacrées à leurs particularités acoustiques et prosodiques restent peu nombreuses. Ce type d'études est motivé par la nécessité d'améliorer l'automatisation de certaines tâches dont le but est d'isoler correctement ces événements dans le flux de parole continu (Shriberg, 1991). Ainsi, dans le cadre de la transcription automatique de la parole, la prise en compte d'un flux acoustique continu implique la modélisation de la totalité des phénomènes présents dans le signal de parole : des mots associés à une représentation phonologique dans le dictionnaire de prononciations, mais également des respirations, des hésitations, des fragments de mots peu ou mal articulés, etc. (Adda-Decker, 2006). Par conséquent, afin d'aboutir à une modélisation fiable des

hésitations vocaliques, il est important de trouver des particularités permettant de les isoler en tant que telles dans le flux de parole et de ne pas les confondre avec d'autres réalisations similaires, notamment des voyelles intralexicales allongées. L'analyse des hésitations vocaliques soulève ainsi la question du marquage au sein d'un ensemble d'items partageant des traits (acoustico-prosodiques) communs, dont les allongements vocaliques de segments intralexicaux, mais également les respirations<sup>1</sup>, etc. Pour ce qui est de l'identification automatique des langues, l'analyse acoustico-prosodique des hésitations vocaliques a comme but de trancher entre deux modèles potentiels du phénomène : un premier, indépendant de la langue, qui suppose que les hésitations vocaliques sont des réalisations « universelles » partageant des traits communs à travers les langues, et un second, dépendant de la langue, qui pose une variabilité interlangues des hésitations vocaliques comprenant des traits acoustico-prosodiques spécifiques. Dans ce dernier cas, les hésitations vocaliques feraient partie des phénomènes spécifiques permettant de caractériser une langue dans un ensemble d'idiomes.

Parmi les paramètres analysés, la fréquence fondamentale ( $F_0$ ), la durée et le timbre (la distribution des voyelles supports<sup>2</sup> des hésitations vocaliques dans le triangle ou quadrilatère vocalique obtenu grâce aux valeurs des deux premiers formants F1 et F2) se sont révélés particulièrement informatifs et ont permis de caractériser ce phénomène dans différentes langues (par exemple, (Shriberg, 1994), pour l'anglais américain, (Swerts, 1998), pour le néerlandais, (Zhao *et al.*, 2005), pour le chinois mandarin, etc.). Les hésitations vocaliques semblent avoir des traits communs dans plusieurs langues en ce qui concerne la durée et la fréquence fondamentale ( $F_0$ ) (Shriberg, 1994 ; Clark *et al.*, 2002 ; Candea, 2000 ; Duez, 1991 ; Zhao *et al.*, 2005 ; Vasilescu *et al.*, 2005), etc. Il s'ensuit des études citées ci-dessus que la durée moyenne des hésitations vocaliques est particulièrement importante comparée à d'autres segments, avoisinant 300 ms dans toutes les langues, bien au-delà de la durée des voyelles intralexicales (estimée autour de 80 ms) et dépassant celle d'une syllabe accentuée (estimée autour de 200 ms), tandis que la  $F_0$  présente un contour plat ou légèrement descendant, éventuellement en rupture avec le contexte lexical environnant. Les motifs « universels » déployés par les deux paramètres durée et  $F_0$  appuient l'hypothèse de la dominante physiologique primant dans la production des hésitations et du statut indépendant de la langue du phénomène. Les deux paramètres semblent être particulièrement robustes pour isoler les hésitations vocaliques dans le flux de parole. Ils sont donc susceptibles d'être utiles dans le cadre de la transcription automatique de la parole.

---

1. Dans le flux de parole continu les respirations apparaissent comme des zones de friction, de durée généralement supérieure à celle d'un segment intralexical et occupant une position interlexicale. Elles partagent ainsi la caractéristique « friction » avec certains segments intralexicaux et la caractéristique « durée importante » et « position interlexicale » avec les hésitations.

2. Nous définissons la voyelle support comme le segment vocalique le plus long et le plus stable d'une hésitation vocalique, possédant ou non un segment adjacent (par exemple, coda nasale).

Pour ce qui est du timbre des hésitations vocaliques, les études lui étant consacrées sont moins nombreuses que celles consacrées à la durée et à la  $F_0$ . La plupart des études présentent le timbre comme un paramètre qui caractérise les hésitations vocaliques de la même façon à travers les langues : la voyelle support de l'hésitation vocalique est plutôt neutre, correspondant à une position relativement centrale de type « schwa » du triangle vocalique, voire proche d'un état de repos articulatoire (Shriberg, 1994 ; Clark *et al.*, 2002). Dans des études publiées récemment dans lesquelles nous décrivons des analyses formantiques de voyelles supports des hésitations vocaliques dans plusieurs langues, nous apportons toutefois des éléments soutenant l'hypothèse que le timbre des hésitations vocaliques dépend des langues (Candea *et al.*, 2005 ; Vasilescu *et al.*, 2006 ; Vasilescu *et al.*, 2007).

Nos travaux portent sur l'analyse des paramètres acoustiques et prosodiques des hésitations vocaliques multilingues dans les grands corpus oraux employés en traitement automatique de la parole et des langues<sup>3</sup>. L'objectif de nos démarches est de caractériser les hésitations vocaliques à travers leurs propriétés acoustiques et prosodiques afin de mettre en valeur les traits universels, c'est-à-dire partagés par toutes les langues *vs.* ceux qui permettent de définir le phénomène au sein d'un idiome. L'intérêt plus large de notre travail est à mettre en lien avec le traitement automatique, et notamment avec l'identification automatique des langues : nous voulons construire des modèles pertinents d'hésitations vocaliques dépendants ou indépendants des langues.

Dans la présente contribution nous nous intéressons aux particularités acoustiques et prosodiques des hésitations vocaliques autonomes dans deux perspectives, inter et intralangue. Nous considérons ces particularités à travers les trois paramètres traditionnellement étudiés pour décrire les hésitations vocaliques, timbre (les deux premiers formants F1 et F2), durée et fréquence fondamentale ( $F_0$ ). Une attention toute particulière sera accordée au timbre des hésitations vocaliques. D'une part, les hésitations de langues différentes seront comparées à travers leurs spécificités acoustiques et prosodiques afin de répondre à la question suivante : déploient-elles un timbre proche d'une réalisation commune, plutôt de type voyelle centrale, ou bien présentent-elles des spécificités dépendantes des langues ? Notre étude porte sur trois langues : anglais américain, français et espagnol. D'autre part, nous aborderons la question des particularités acoustiques et prosodiques des hésitations vocaliques en rapport avec le système (vocalique) de chaque langue. À l'issue d'une analyse interlangue du timbre, nous serons ainsi en mesure de situer le timbre des hésitations par rapport à l'inventaire vocalique de la langue.

---

3. Pour une bibliographie complète nous renvoyons le lecteur à : <http://www.limsi.fr/individu/ioana/>.

## 2. Corpus et méthodologie

Nous faisons appel ici à des données orales présentes dans notre laboratoire grâce à différents projets européens et internationaux : il s'agit essentiellement de données enregistrées et transcrites de type « parole semi-préparée » (journaux télévisés, débats parlementaires). Ainsi, environ 20 heures d'émissions radiotélévisées de sources diverses ont été utilisées pour le français (France Inter, France Info, France 2,...). Dix heures de données comparables ont fourni les hésitations vocaliques en anglais américain (les sources étant CNN, VOA, ABC, etc.) et également 10 heures d'enregistrements de débats du Parlement européen ont été utilisées pour extraire les hésitations en espagnol. Les émissions radiotélévisées et les débats parlementaires illustrent le style de parole appelé « discours public » et caractérisé par la préplanification, car, dans les deux cas, les locuteurs construisent leurs interventions sur la base de notes écrites (O'Connell *et al.*, 2004). Le discours public est essentiellement constitué de monologues, même si le dialogue n'est pas entièrement exclu (par exemple, les interviews). Lorsqu'elles apparaissent dans un monologue, les hésitations vocaliques ont le rôle de contribuer à la structuration du discours (Swerts, 1997). Dans un dialogue, les hésitations vocaliques sont plus nombreuses car elles sont la principale marque de l'oralité, de la planification du discours et de la régulation des tours de parole. Par conséquent, nous nous attendons à trouver moins d'hésitations dans nos données que dans une parole « spontanée ». Il est également possible que les hésitations soient relativement prototypiques car témoignant d'une même fonction, celle des marques de la structuration du discours. Enfin, les émissions radiotélévisées ainsi que les débats parlementaires sont destinés à une large audience : nous nous attendons aussi à ce que les hésitations correspondent à une prononciation relativement soignée.

Pour cette étude, entre 1 000 et 4 000 occurrences d'hésitations vocaliques, en fonction des langues, prononcées par quelques dizaines de locuteurs par langue, ont été extraites automatiquement<sup>4</sup> grâce au système d'alignement automatique développé au LIMSI (Gauvain *et al.*, 2002). Les hésitations vocaliques ont également été vérifiées manuellement afin d'éliminer des erreurs potentielles d'extraction (par exemple, des hésitations confondues avec de la musique, de fortes respirations ou d'autres bruits). Simultanément des voyelles intralexicales dans tous les contextes lexicaux attestés ont été extraites automatiquement. Les analyses se basent essentiellement sur les productions de locuteurs masculins, plus nombreux dans les données journalistiques et les débats parlementaires.

L'analyse des paramètres retenus pour cette étude (fréquence fondamentale ( $F_0$ ), durée, timbre à travers les deux premiers formants F1 et F2) a été menée avec le

---

4. Nous avons voulu retenir pour cet article des hésitations prototypiques notamment en terme de timbre, donc délimitées par des frontières temporelles du contexte lexical environnant afin d'exclure toute confusion avec des voyelles intralexicales (voir sections suivantes). À cet effet nous avons extrait des hésitations vocaliques autonomes (par exemple, en français des « euh » précédés et suivis d'une pause silencieuse).

logiciel Praat (Boersma *et al.*, 1999). Une valeur a été extraite toutes les 5 ms pour la fréquence fondamentale et les formants. Pour chaque segment correspondant à une hésitation vocalique autonome, un taux de voisement a été calculé en tant que rapport entre le nombre de trames voisées et le nombre total de trames. Nous avons retenu uniquement les segments dont le taux de voisement était supérieur à 40 %. Cette démarche a exclu 27 % des occurrences en anglais américain, 23 % en français et 10 % en espagnol. Ce choix a été fait afin de conserver les exemplaires les plus prototypiques compte tenu des paramètres objectifs mesurés.

Pour chaque segment retenu, nous avons calculé des valeurs moyennes pour les trois fréquences (fréquence fondamentale et formants F1 et F2) comme moyenne sur les trois trames voisées centrales du segment.

### 3. Timbre

Dans cette section nous allons nous intéresser au timbre des hésitations vocaliques à travers les langues (analyse interlangue) et en rapport avec l'inventaire vocalique d'une langue donnée (intra-langue) en mesurant les deux premiers formants F1 et F2 qui permettent de positionner le segment vocalique dans le triangle ou le quadrilatère vocalique.

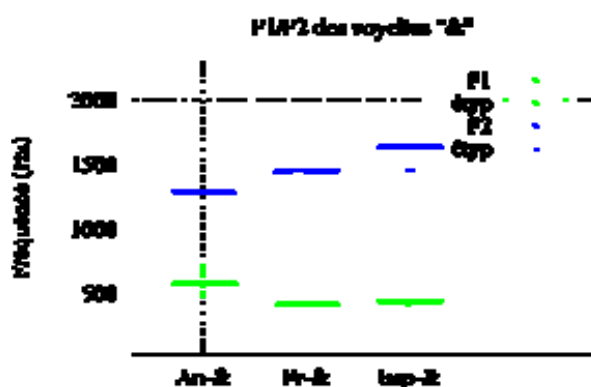
#### 3.1. *Timbres des hésitations vocaliques à travers les langues*

Dans l'introduction nous avons mentionné les différentes transcriptions orthographiques dont bénéficient les hésitations à travers les langues, par exemple, « uh/um » en anglais américain, « euh » en français ou bien « eh » en espagnol. Ces différentes transcriptions laissent supposer que les auditeurs natifs associent les hésitations vocaliques à des voyelles différentes selon les langues.

Nous avons procédé à l'analyse comparative des timbres des hésitations dans les trois langues analysées à travers des mesures des deux premiers formants F1 et F2. Le but de cette analyse est de voir si les transcriptions orthographiques différentes peuvent être mises en lien avec des différences objectives de qualité vocalique. La figure 1, qui représente des valeurs moyennes F1 et F2 par voyelle d'hésitation et par langue, suggère des différences entre les trois langues. La voyelle d'hésitation est encodée avec le symbole « & »<sup>5</sup>.

---

5. Nous encodons les hésitations vocaliques avec le symbole indépendant de la langue « & » : nous allons employer ce symbole dans les figures de cet article lorsqu'il s'agit des hésitations vocaliques en général. Nous utiliserons les transcriptions orthographiques lorsque nous ferons référence à l'hésitation vocalique d'une langue donnée.



**Figure 1.** Moyennes et écarts types des premiers deux formants F1 et F2 des voyelles d'hésitation (« & ») en anglais américain, français et espagnol

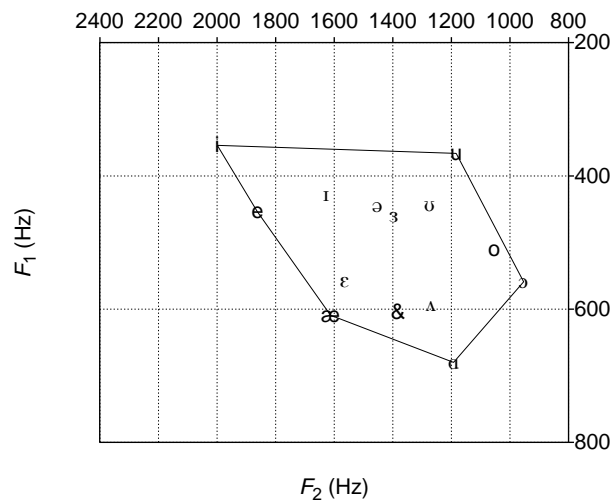
Les valeurs moyennes obtenues pour les voyelles d'hésitation en anglais américain, français et espagnol soutiennent l'hypothèse d'un timbre dépendant de la langue. L'hypothèse du timbre central (ou « neutre », de type « schwa ») apparaît ainsi contestable : nos données montrent que le timbre est ouvert central en anglais américain, mi-fermé antérieur en espagnol et mi-fermé central ou antérieur de type /œ/ en français. Les valeurs des formants F1 et F2 obtenues sont donc en accord avec les transcriptions orthographiques mais aussi avec la perception des hésitations par les auditeurs (Vasilescu *et al.*, 2005). Afin de vérifier si ces différences sont significatives, une analyse de la variance a été menée (MANOVA, valeur de F et niveau de significativité estimés par le Lambda de Wilks). Cette analyse a été utilisée pour tester l'effet du facteur « langue » sur les deux premiers formants F1 et F2 considérés ensemble. Les différences sont statistiquement significatives ( $F(4,436) = 79,96$ ,  $p < 0,0001$ ). Le timbre de la voyelle d'hésitation en anglais américain s'avère être le plus ouvert et le plus postérieur, en comparaison avec le français et l'espagnol dont les hésitations sont d'aperture similaire, mi-fermée.

Par conséquent, les données analysées ici viennent à l'appui de l'hypothèse d'un timbre dépendant de la langue des hésitations, lequel ne privilégierait pas ainsi – et par défaut – la zone centrale du triangle/quadrilatère vocalique comme il a été avancé par H. Clark et J. Fox Tree (Clark *et al.*, 2002).

### 3.2. Timbre des voyelles d'hésitation vs. voyelles intralexicales

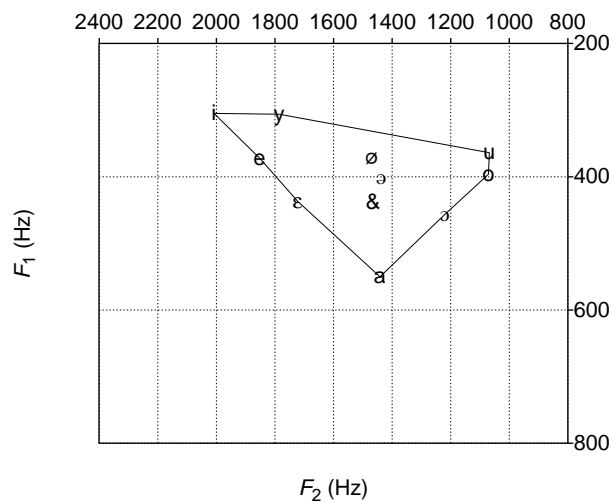
L'analyse interlangue du timbre des voyelles d'hésitation a montré que leur timbre varie d'une langue à l'autre. Dans cette section nous considérons le timbre de la voyelle d'hésitation en rapport avec le système de la langue, *i.e.*, l'inventaire vocalique. Pour cette comparaison, nous avons utilisé des données formantiques (F1

et F2) extraites avec la stratégie automatique décrite en section 2 et pour chaque langue<sup>6</sup>. Ensuite, des valeurs moyennes par voyelle et par langue ont été calculées. La figure 3 montre le système vocalique obtenu pour l'anglais américain en utilisant les valeurs moyennes par segment vocalique placées dans l'espace déterminé par les deux premiers formants F1 et F2. La voyelle d'hésitation (encodée « & ») est, elle aussi, représentée dans ce même espace F1 et F2. Cette représentation permet de visualiser une distance objective entre voyelle d'hésitation et voyelles intralexicales de l'inventaire vocalique de l'anglais américain. Les figures 3 et 4 montrent la distribution des voyelles d'hésitation et intralexicales dans l'espace F1 et F2 pour le français et l'espagnol respectivement.

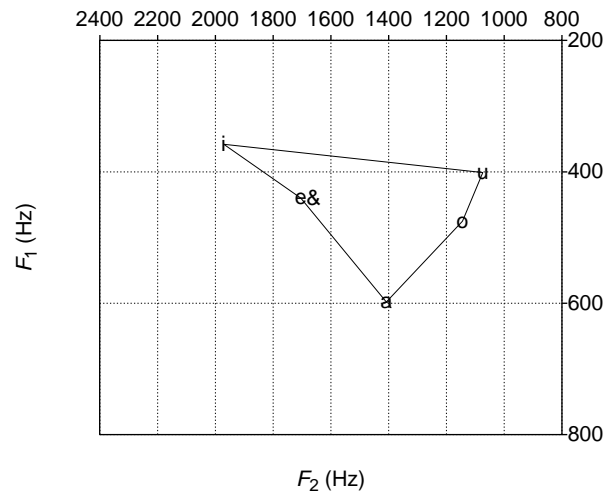


**Figure 2.** Valeurs moyennes F1 et F2 des voyelles d'hésitation et intralexicales en anglais américain

6. Les données utilisées nous ont fourni 270 k voyelles en anglais américain, 120 k en français et 140 k en espagnol. La distribution de ces données par timbre est détaillée dans le tableau 2, consacré à la fréquence des différents timbres vocaliques.



**Figure 3.** Valeurs moyennes F1 et F2 des voyelles d'hésitation et intralexicales en français



**Figure 4.** Valeurs moyennes F1 et F2 des voyelles d'hésitation et intralexicales en espagnol

La comparaison entre les timbres respectifs des voyelles d'hésitation et des voyelles intralexicales montre que ces premières sont proches ou presque identiques à certains timbres attestés dans l'inventaire vocalique des langues analysées. En

cherchant pour chaque langue les plus proches voisins de « & », nous trouvons une distance intervocalique minimale entre la voyelle d'hésitation et des timbres intralexicaux /&/ vs. /ə/ (/ə/+/œ/) en français<sup>7</sup> et /&/ vs. /e/ en espagnol. Pour ce qui est de l'anglais américain, la voyelle d'hésitation se positionne dans une zone d'aperture ouverte sur l'axe central du quadrilatère vocalique sans pour autant privilégier une proximité particulière avec un timbre intralexical. Les voisins les plus proches sont : /ʌ/ et /æ/.

Afin d'évaluer la proximité de qualité vocalique entre la voyelle d'hésitation et les voyelles intralexicales pour chaque langue considérée ici, nous avons mené des analyses statistiques MANOVA. L'effet du « type de voyelle » (la voyelle d'hésitation vs. les voyelles intralexicales) a été ainsi testé sur les deux formants F1 et F2 considérés ensemble. En français, la voyelle d'hésitation a été comparée au timbre /ə/ (comme défini plus bas). Le résultat a montré que la différence entre les deux types de voyelles est statistiquement significative (Lambda de Wilks,  $F(2, 2905) = 291,64$ ,  $p < 0,0001$ ). La voyelle d'hésitation s'avère significativement plus ouverte (F1,  $p < 0,0001$ ). Elle est également plus antérieure que sa contrepartie intralexicale (F2,  $p < 0,001$ ). La voyelle d'hésitation espagnole se confond à l'écoute avec son homologue intralexical /e/. Un effet du type de voyelle a été pourtant observé (MANOVA, Lambda de Wilks,  $F(2, 2931) = 285,74$ ,  $p < 0,0001$ ). Elle est plus ouverte (F1) et surtout plus postérieure (F2) que le segment intralexical de timbre proche /e/ ( $p < 0,001$ ). Le caractère ouvert des voyelles d'hésitation pourrait être mis en relation avec la durée importante de ces segments (voir section suivante), tandis que leur centralisation sur l'axe F1 et surtout sur F2 peut éventuellement être une conséquence de la « pression » exercée par la « position neutre » (centrale). Enfin, un effet statistiquement significatif du « type de voyelle » peut également être observé en anglais américain (MANOVA, Lambda de Wilks,  $F(4,480) = 35,48$ ,  $p < 0,0001$ ). Les différences entre la voyelle d'hésitation et les voyelles intralexicales de timbre proche concernent l'axe antérieur/postérieur (F2). La voyelle d'hésitation représente un timbre intermédiaire entre les deux segments phonémiques /ʌ/ et /æ/ et cette différence est statistiquement significative ( $p < 0,001$ ).

À l'issue de cette section, il est à noter la variabilité interlangue du rapport entre une voyelle d'hésitation et un inventaire vocalique phonémique. Les timbres moyens des voyelles d'hésitation sont proches d'une position plutôt centrale, mi-ouverte en français et plutôt ouverte en anglais américain, tandis qu'en espagnol la voyelle d'hésitation approche un timbre antérieur d'aperture moyenne.

---

7. Dans le cadre du traitement automatique nous rassemblons sous l'étiquette « 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. Cette voyelle réunissant les deux timbres /ə/+/œ/ est encodée ici avec le symbole « schwa » (/ə/). En effet, lorsque les « schwas » sont réalisés, leur timbre est très proche, voire identique à [œ], [ø] (Fougeron, Gendrot et Bürki, 2007).

### 3.3. Discussion

L'analyse du timbre des hésitations vocaliques en anglais américain, français et espagnol, a mis en évidence des spécificités inter (fréquences formantiques absolues) et intralangue (position relative au sein du système vocalique). Tout d'abord, les hésitations vocaliques présentent des timbres moyens différents d'une langue à l'autre. Nos données soutiennent l'hypothèse d'un timbre dépendant de la langue, lequel, de surcroît, ne privilégie pas forcément une position de la voyelle d'hésitation par défaut centrale dans l'espace vocalique ou encore « neutre », proche de « schwa ». Les données présentées ici montrent que la voyelle d'hésitation accepte des réalisations différentes dans les trois idiomes pris en compte, allant d'une voyelle centrale d'aperture moyenne ou plus ouverte en français et anglais américain respectivement, à un timbre plutôt proche de la voyelle /e/, donc antérieur en espagnol. Ainsi, les voyelles d'hésitation s'avèrent proches de timbres attestés dans le système occupant plutôt une aire centrale/antérieure et d'aperture variable sur le continuum ouvert/mi-ouvert mais ne coïncident pas avec leurs contreparties intralexicales. Enfin, les différences observées entre voyelles d'hésitation de différentes langues et voyelles d'hésitation *vs.* intralexicales ont une significativité statistique.

À l'issue de cette section, nous nous sommes interrogés sur la préférence des voyelles d'hésitation pour certaines qualités vocaliques. La présente étude porte sur trois langues européennes : la préférence de ces trois langues pour des qualités vocaliques centrales et/ou antérieures ne permet pas une généralisation à toutes les langues possédant des hésitations vocaliques. Cependant, d'autres études consacrées aux hésitations vocaliques laissent supposer des spécificités de timbre et fournissent des appuis supplémentaires à notre analyse. Le tableau 1 répertorie les différentes transcriptions orthographiques des hésitations vocaliques dans sept autres langues, issues de familles linguistiques différentes. Les transcriptions orthographiques suggèrent que les timbres des voyelles d'hésitation de ces langues privilégient, eux aussi, une zone antéro-centrale d'aperture moyenne. S'agit-il d'une préférence répandue pour ce type de qualité vocalique ou bien les hésitations vocaliques pourraient-elles *a priori* occuper toute position du triangle/quadrilatère vocalique ? Les timbres postérieurs sont-ils exclus des timbres supports des hésitations vocaliques ? Nos données et les données disponibles jusqu'à présent semblent appuyer cette hypothèse : quelles seraient alors les causes possibles de cette préférence ?

Langues	Hésitations vocaliques (transc. ortho.)	Auteurs
Suédois	eh, öh	(Eklund <i>et al.</i> , 1998)
Allemand	äh, ähm, häs	(Batliner <i>et al.</i> , 1995)
Néerlandais	eeh	(Van Donzel <i>et al.</i> , 1998)
Norvégien	e, eh, em, øhø, aj	(Svennevig, 1999) (cité par (Clark <i>et al.</i> , 2002))
Japonais	e	(Watanabe <i>et al.</i> , 2006)
Chinois mandarin	uh, um (conv. anglaises)	(Zhao <i>et al.</i> , 2005)
Hébreu	eh, em, ah, am	(Maschler, 1997) (cité par (Clark <i>et al.</i> , 2002))

**Tableau 1.** Transcriptions orthographiques des hésitations vocaliques dans différentes langues

Les travaux consacrés à l'organisation des espaces vocaliques montrent la préférence des langues pour les voyelles antérieures (Lindblom, 1986). Dans la théorie de la dispersion de Lindblom ainsi que dans celle de la dispersion-focalisation de (Schwartz *et al.*, 1997), il a été montré que la distance entre les voyelles [i] et [a] est supérieure à la distance entre [i] et [u] laissant plus de possibilités pour les langues à l'avant du triangle ou quadrilatère vocalique plutôt qu'à l'arrière. L'espace vocalique maximal généré par (Boë *et al.*, 1989) montre aussi cette tendance. Cette tendance est également présente dans les espaces vocaliques des trois langues analysées ici (notamment les figures 6, 7 et 8 plus bas). Dans le cadre de cette théorie les hésitations vocaliques se placeraient ainsi dans la région de l'espace vocalique la plus exploitée. D'autre part, l'analyse de la fréquence des phonèmes vocaliques à partir de données phonologiques issues d'un échantillon représentatif des langues du monde confirme cette tendance : la voyelle centrale ouverte [a] et la voyelle antérieure fermée [i] sont les plus fréquentes dans les langues (Marsico *et al.*, 2004).

La voyelle d'hésitation n'a pas un statut phonologique. Cependant, elle semble être « attirée » par des timbres phonémiques attestés dans le système vocalique d'une langue. En accord avec la théorie mentionnée ci-dessus, ces timbres sont prédictibles : il s'agit des phonèmes de la zone antéro-centrale de l'espace vocalique, particulièrement exploitée et dont certains timbres comme [i] et [a] sont très fréquents dans les langues du monde.

Nous avons également vérifié si nos données orales en général pouvaient être mises en relation avec les travaux sur l'organisation des espaces vocaliques et la

fréquence de certains phonèmes vocaliques. À cet effet, nous avons calculé le pourcentage d'items vocaliques dans le corpus par timbre et par langue. Les résultats sont montrés dans le tableau 2. Nos données montrent que les langues considérées ici ont une préférence certaine pour les voyelles ouvertes centrales et antérieures qui s'avèrent les mieux représentées dans les trois inventaires vocaliques (entre 10 et 30 % d'items vocaliques/langue). En anglais américain, le timbre central /ə/ est le mieux représenté (21,5 %), suivi par /ɪ/ (16,1 %), /i/ (11,3 %), /ʌ/ (4,8 %) et /æ/ (8 %). En français, un timbre central ouvert est également le plus fréquent, /a/ (21,8 %), suivi par des timbres antérieurs, /i/ (15,5 %), /e/ (16,1 %) et /ɛ/ (13,5 %). Enfin, en espagnol, /e/ est de loin le timbre le plus fréquent dans nos données (30,7 %), suivi par /a/ (24,5 %) et /i/ (16,6 %). L'espagnol fournit toutefois une exception : la voyelle postérieure d'aperture moyenne /o/ avec 21,5 % des données. L'attraction de la voyelle d'hésitation par le timbre phonémique le plus fréquent se vérifie ainsi en espagnol. En anglais américain et en français elle n'est pas attirée par les qualités vocaliques fréquentes, laissant supposer une force d'attraction plus importante de la position centrale, « neutre ».

Langues	Voyelles	Nbr. d'items (k)	%
Anglais am.	/i/	3	11,3
	/ɪ/	4	16,1
	/ɛ/	2	8,4
	/e/	1,5	5,8
	/æ/	2	8
	/ɑ/	1	4,2
	/ʌ/	1	4,8
	/ɔ/	1	4,8
	/o/	1	3,7
	/u/	1	4,2
	/ʊ/	0,3	1,3
	/ə/	6	21,5
	/ɜ/	1,5	6
Français	/i/	13	15,5
	/e/	14	16,1
	/ɛ/	11	13,5
	/a/	18	21,8
	/ɔ/	5	6,2
	/o/	3	3,3
	/u/	4	5,3
	/ø/	1	1,3
	/y/	5	5,4
	/ə/ (/œ/ + /ə/)	10	11,8
Espagnol	/i/	67	16,6
	/e/	124	30,7
	/a/	100	25
	/o/	87	21,5
	/u/	26	6,5

Tableau 2. Fréquence des voyelles intralexicales

Ces résultats suggèrent que l'hypothèse formulée au sujet de la fréquence de certains phonèmes à travers les langues se vérifie lorsqu'il s'agit de segments vocaliques extraits automatiquement à partir de grands corpus oraux : les langues analysées ici ont une préférence pour les voyelles centrales et antérieures, lesquelles sont les mieux représentées dans le lexique. Nos données apportent ainsi un appui supplémentaire aux travaux sur l'organisation des espaces vocaliques et sur la fréquence de certains timbres vocaliques.

À partir de ces observations, nous avançons ici l'hypothèse d'une préférence des hésitations pour certaines qualités vocaliques comme caractérisées plus haut. Cette hypothèse n'exclurait pas le rôle potentiel « d'attracteur » du timbre central ou « neutre ». Cependant, il est envisageable que cette position, associée généralement à un timbre de type « schwa », varie à travers les langues (Gick *et al.*, 2004). Le caractère universel ou dépendant des langues des bases articulatoires (articulatory settings) reste encore à prouver. Cependant des résultats expérimentaux à l'issue d'une étude consistant en l'exploration de la position des articulateurs en repos ou en articulation avec des rayons X que B. Gick et ses collaborateurs ont récemment menée semblent conforter l'hypothèse de la dépendance de la langue. Par extension, si cette hypothèse est confirmée, les hésitations vocaliques associées souvent à un état proche du repos des articulateurs, auraient, elles aussi, des cibles articulatoires différentes d'une langue à l'autre. Cette variabilité entraînerait de légères différences de timbre, des hésitations préférant une position centrale. Le français et l'anglais américain semblent conforter cette tendance, les voyelles d'hésitation de ces deux langues occupant une zone plutôt centrale du triangle ou quadrilatère vocalique. Le timbre de la voyelle d'hésitation en espagnol est le plus extrême sur l'axe antéro-postérieur. Cette langue montre une certaine résistivité à exploiter le centre de l'espace vocalique : le timbre phonémique correspondant s'avère être un « attracteur » plus fort. Nos données montrent que cette qualité vocalique phonémique est également la plus représentée dans le corpus (30 % des occurrences, cf. tableau 2).

Par conséquent, le timbre des hésitations vocaliques pourrait être déterminé par des contraintes antagonistes : d'une part, la position de repos « attirerait » la voyelle d'hésitation vers le centre de l'espace vocalique ; d'autre part, des contraintes acoustiques et informationnelles (organisation des espaces vocaliques et fréquence lexicale des voyelles dans les langues) rapprocheraient le timbre des voyelles d'hésitation de certains timbres phonémiques. Cette hypothèse mérite d'être étayée par des analyses supplémentaires. Par exemple, une approche typologique (*i.e.*, la comparaison avec d'autres espaces vocaliques) pourrait apporter des éléments permettant de généraliser ces observations. Une perspective diachronique sur la stabilité de la voyelle d'hésitation et sur son comportement compte tenu de l'évolution dans le temps d'un système vocalique pourrait également être utile.

#### 4. Durée

Les hésitations vocaliques sont décrites comme étant significativement plus longues que leurs homologues intralexicales, aussi bien du point de vue objectif que perceptif (Candea *et al.*, 2005 ; Duez, 1991 ; Shriberg, 1994 et 2001). Leur durée peut également être mise en lien avec le style de parole et l'état affectif du locuteur : dans la parole non préparée et lorsque les conditions d'élocution sont susceptibles d'engendrer du stress, les hésitations vocaliques sont plus fréquentes et plus longues (Vasilescu *et al.*, 2006). Grâce à l'alignement automatique des données orales employées pour la présente étude, nous avons pu évaluer la durée des hésitations vocaliques en anglais américain, français et espagnol et comparer la durée des voyelles d'hésitation *vs.* intralexicales.

##### 4.1. Durée des voyelles d'hésitation à travers les langues

Des études menées sur différentes langues ont mis en évidence des durées de voyelles d'hésitation très importantes comparées à leurs contreparties phonémiques. Les hésitations atteignent une valeur moyenne de 300 ms, par exemple : (Clark *et al.*, 2002), (Brennan *et al.*, 1995) pour l'anglais américain, (Swerts, 1998) pour le néerlandais, (Zhao *et al.*, 2005) pour le chinois mandarin, (Candea, 2000) pour le français, etc. Pour ce qui est des langues qui possèdent deux hésitations vocaliques, une première, pleinement vocalique, et une seconde, de type voyelle + coda nasale comme l'anglais américain ou le néerlandais, ces langues montrent des différences de durée entre les deux hésitations. En effet, les hésitations de type voyelle + coda nasale s'avèrent être plus longues, probablement pour des raisons physiologiques (Clark *et al.*, 2002). Dans nos données en anglais américain, l'hésitation pleinement vocalique « uh » représente environ 85 % des données.

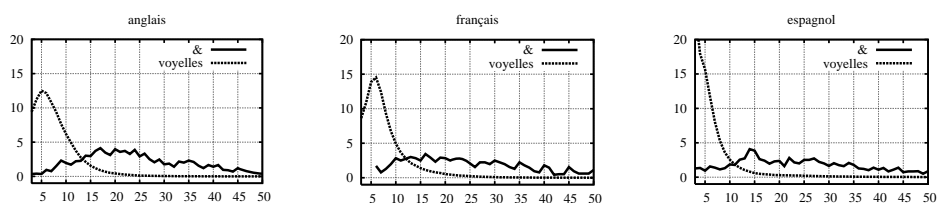
Le tableau 3 montre les durées moyennes des hésitations vocaliques dans les trois langues analysées ici. Les données en anglais américain sont divisées selon les types d'hésitations possibles « uh » ou « um ». Nos données confirment les observations antérieures quant à la durée moyenne des hésitations vocaliques : la durée environne 300 ms et s'avère être un paramètre à la fois indépendant de la langue et du corpus. Une variabilité importante au sein de chaque langue est également à noter, même si la plupart des hésitations vocaliques dépassent 250 ms.

	Anglais am.	Français	Espagnol
<b>Nbr. échant.</b>	3 136/580	1 199	1 993
<b>Moyenne</b>	322 <i>uh</i> / 423 <i>um</i>	323	315
<b>Écart type</b>	172 / 189	201	207
<b>Min</b>	60 / 60	60	60
<b>Max</b>	1 570 / 1 350	1 740	1 390
<b>Médiane</b>	280 / 400	280	270

**Tableau 3.** Durées (ms) des voyelles d'hésitation en anglais américain, français et espagnol : moyenne, écart type, min, max, médiane

#### 4.2. Durée des voyelles d'hésitation vs. voyelles intralexicales

La figure 5 montre la distribution des durées pour les segments vocaliques d'hésitation vs. intralexicaux et pour chaque langue. Tandis que les voyelles phonémiques atteignent une valeur modale se situant autour de 60 ms (voire plus courte en espagnol), les hésitations montrent des distributions plutôt homogènes avec une concentration de valeurs plus importante entre 150 à 300 ms. Les segments intralexicaux vont rarement au-delà de 150 ms. Les voyelles d'hésitation, quant à elles, durent au moins deux fois plus qu'un segment intralexical de durée moyenne.



**Figure 5.** Distribution de la durée des segments d'hésitation et intralexicaux en anglais américain, français et espagnol. Axe X : durée des segments (cs). Axe Y : % des segments

Le tableau 4 montre les durées moyennes des voyelles d'hésitation et phonémiques. Nos données confirment ainsi des observations antérieures : les voyelles d'hésitation sont en effet significativement plus longues que leurs homologues phonémiques (ANOVA à un facteur,  $p < 0,001^8$ ).

8. Afin que les données soient adaptées aux hypothèses de l'ANOVA, et notamment à celle de la distribution normale, le calcul a été effectué sur les logarithmes de durée.

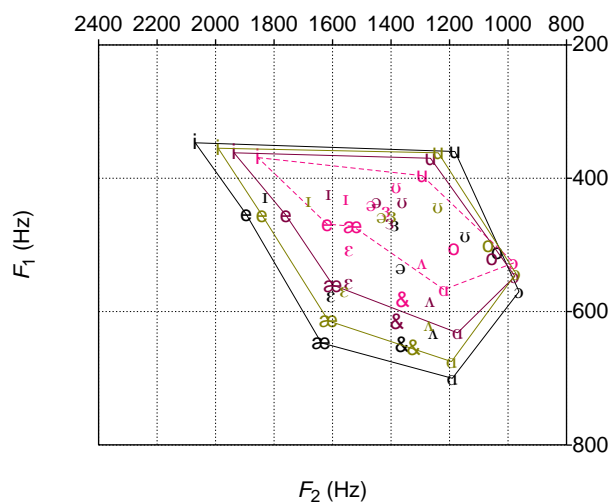
Langue	Voyelle	Durée (ms)
Anglais am.	/i/	93
	/I/	59
	/ɛ/	75
	/e/	119
	/æ/	105
	/ɑ/	107
	/ʌ/	72
	/ɔ/	115
	/o/	122
	/u/	95
	/ʊ/	70
	/ə/	49
	/ɜ/	87
	<b>/&amp;/</b>	<b>322</b>
Français	/i/	93
	/e/	82
	/ɛ/	70
	/a/	70
	/ɔ/	63
	/o/	102
	/u/	74
	/ø/	90
	/y/	77
	/ə/ (/œ/ + /ə/)	87
	<b>/&amp;/</b>	<b>323</b>
Espagnol	/i/	64
	/e/	58
	/a/	71
	/o/	71
	/u/	60
	<b>/&amp;/</b>	<b>315</b>

**Tableau 4.** *Durée moyenne des voyelles d'hésitation vs. intralexicales*

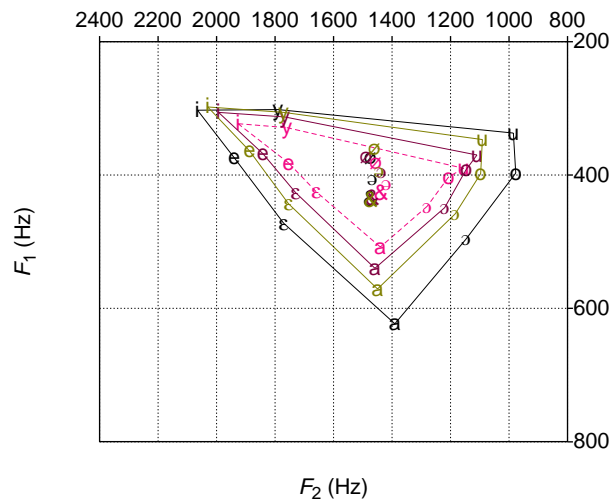
### 4.3. Effet de la variation de la durée sur le timbre de la voyelle d'hésitation vs. voyelles intralexicales

(Gendrot *et al.*, 2005) ont montré que les systèmes vocaliques tendent à occuper une position plus centrale dans l'espace déterminé par les deux premiers formants  $F_1$  et  $F_2$  avec la réduction de la durée des segments. L'espace vocalique augmente avec la durée de ses segments, les voyelles périphériques devenant ainsi de plus en plus périphériques. Dans la section précédente, nous avons noté une distribution de la durée différente pour les segments d'hésitation vs. phonémiques. Nous avons également observé un comportement dépendant de la langue des hésitations en terme de qualité vocalique : les voyelles d'hésitation des trois langues analysées occupent une aire centrale/antérieure dans l'espace vocalique et leurs timbres moyens s'avèrent proches de timbres par ailleurs attestés dans l'inventaire phonémique de la langue.

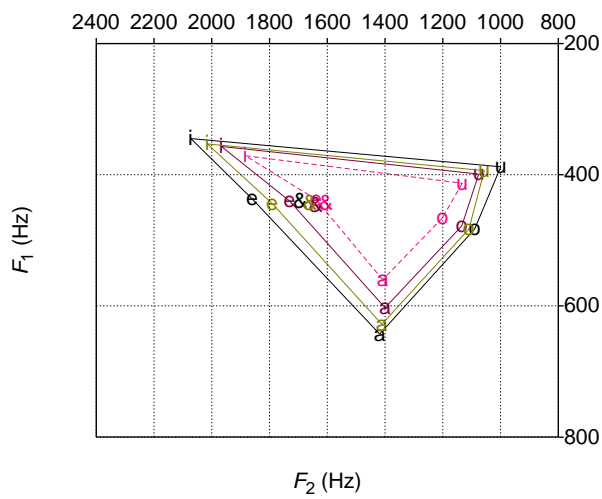
Afin de mieux circonscrire le rapport entre le comportement du timbre des voyelles d'hésitation vs. système de la langue, nous comparons ici l'effet de la durée sur la voyelle d'hésitation comparé au même effet sur les segments de l'inventaire vocalique phonémique de la langue et ce pour les trois idiomes considérés dans cette étude, anglais américain, français et espagnol. Plus précisément, nous allons discuter la question de la centralisation des segments vocaliques d'hésitation avec la réduction de la durée. Présentent-ils un comportement comparable à celui observé par (C. Gendrot *et al.*, 2005) pour les voyelles intralexicales ? Ou bien montrent-ils des comportements différents, ce qui soutiendrait un statut extrasystémique des voyelles d'hésitation ?



**Figure 6.** Moyenne  $F_1$  et  $F_2$  pour les voyelles d'hésitation et intralexicales en fonction de la durée en anglais américain



**Figure 7.** Moyenne  $F_1$  et  $F_2$  pour les voyelles d'hésitation et intralexicales en fonction de la durée en français



**Figure 8.** Moyenne  $F_1$  et  $F_2$  pour les voyelles d'hésitation et intralexicales en fonction de la durée en espagnol

À cet effet, les données ont été divisées en quatre paquets de durées, et les valeurs moyennes des deux premiers formants F1 et F2 ont été calculées pour les voyelles d'hésitation correspondant à chaque paquet de durées (tableau 5). Les figures 6, 7 et 8 montrent les systèmes vocaliques et les voyelles d'hésitation de l'anglais américain, du français et de l'espagnol, obtenus selon le critère des différences de durées des segments. Ils confirment la tendance à la centralisation avec la réduction de la durée observée dans (Gendrot *et al.*, 2005) : les espaces concentriques correspondent, à la périphérie, aux durées importantes et au centre aux segments les plus courts. Les différentes voyelles correspondant à des valeurs moyennes de formants F1 et F2, occupent l'espace vocalique d'une manière organisée. La centralisation ne se réalise pas de la même façon d'une langue à l'autre, les segments décrivent des mouvements spécifiques vers le centre de l'espace vocalique selon le timbre et la langue. Néanmoins, toutes les voyelles intralexicales sont concernées par ce mouvement.

Pourcentage de segments selon la durée					
	Type	XCourt	Court	Moyen	Longues
Ang. am.	V	24	21	22	33
	H	12	22	42	22
Français	V	27	30	24	19
	H	12	14	34	40
Espagnol	V	42	27	17	13
	H	15	20	32	33

**Tableau 5.** Distribution du pourcentage de voyelles intralexicales **V** et d'hésitation **H** en anglais américain, français et espagnol selon la durée. **XCourt** : 30-40 ms pour **V**, 30-120 ms pour **H** ; **Court** : 50-60 ms pour **V**, 130-190 ms pour **H** ; **Moyen** : 70-90 ms pour **V**, 200-340 ms pour **H** ; **Longues** : 100 + pour **V**, 350 + pour **H**<sup>9</sup>

Les voyelles d'hésitation affichent des comportements comparables aux voyelles intralexicales. En français, la voyelle d'hésitation, comme la voyelle phonémique

9. Nous avons choisi ces différents seuils de durées afin de contrôler des effets de corpus potentiels, notamment pour ce qui est des segments intralexicaux qui n'ont pas bénéficié d'une évaluation manuelle. Ainsi la classe des segments trop courts (XC) serait la plus susceptible de contenir des erreurs de détection automatique, tandis que les classes C, M pour les V et M, L pour les H correspondraient aux exemplaires les plus prototypiques. La distribution dans les différentes classes des voyelles intralexicales ou d'hésitation selon les langues fournit également des informations sur le débit : par exemple, en espagnol il y a plus de voyelles dans les classes XC et C que dans les autres langues, laissant supposer un débit plus rapide.

proche de /ə/ est relativement stable car elle se trouve au centre de l'espace vocalique. En espagnol, le mouvement centripète concerne l'axe F2, et le comportement de la voyelle d'hésitation est similaire à celui de la voyelle intralexicale /e/. La voyelle d'hésitation décrit aussi un mouvement de centralisation plus important que la voyelle /e/. Cependant, elle n'atteint pas le centre du triangle vocalique lorsque les durées atteignent les valeurs les plus réduites. Enfin, la voyelle de l'anglais américain décrit le mouvement centripète le plus important. La trajectoire est même très similaire à celle des voyelles intralexicales d'aperture maximale et pourrait être mise en lien avec son caractère plus ouvert : comme pour les segments phonémiques, plus la voyelle est périphérique, plus le mouvement vers le centre avec la diminution de la durée est marqué.

C. Gendrot et M. Adda-Decker ont observé un effet statistiquement significatif de la durée sur la centralisation des voyelles phonémiques (Gendrot *et al.*, 2005). Nous avons évalué le même effet sur les voyelles d'hésitation : l'effet observé, c'est-à-dire la centralisation avec la réduction de la durée est présent sans pour autant être statistiquement significatif. Le mouvement centripète moins important pourrait être mis en relation avec la durée des hésitations vocaliques. En effet, les hésitations vocaliques atteignent rarement des durées inférieures à 100 ms, contrairement à leurs homologues phonémiques<sup>10</sup>.

#### 4.4. Discussion

L'analyse inter et intralangue de la durée des hésitations vocaliques a confirmé les observations précédentes à propos de la différence significative de durée entre voyelles d'hésitation et voyelles intralexicales. En effet, les voyelles d'hésitation dans les trois langues analysées atteignent des valeurs moyennes d'environ 300 ms, bien au-delà de la durée maximale d'une voyelle phonémique. L'analyse de l'effet de la variation de la durée sur la centralisation de l'espace vocalique permet de considérer les voyelles d'hésitation comme une interpolation entre une voyelle intralexicale spécifique (par exemple, /e/ en espagnol) et une aire centrale de l'espace vocalique pouvant être associée au timbre « neutre ». Les présents résultats confirment ainsi le caractère dépendant de la langue des hésitations vocaliques.

---

10. Les figures 6, 7 et 8 pourraient donner une image biaisée du phénomène de réduction vocalique : en effet, il serait possible d'imaginer que plus les voyelles intralexicales sont brèves plus les voyelles d'hésitation sont également brèves. Cependant, les données présentées ici ne permettent pas d'établir une corrélation entre les paliers de durées et les types de segments analysés (*i.e.*, hésitations et voyelles intralexicales). En effet, les contextes d'extraction ne sont pas contrôlés à ce stade d'analyse, *i.e.*, les hésitations brèves ne proviennent pas exclusivement de contextes à débit rapide entraînant une réduction vocalique. Il serait sans doute intéressant de prendre en compte le rôle de la variation du débit sur la durée des hésitations vocaliques, comparée à celle des voyelles intralexicales.

## 5. Fréquence fondamentale

La fréquence fondamentale ( $F_0$ ) des voyelles d'hésitation est décrite comme un paramètre indépendant de la langue : les hésitations vocaliques présentent un contour stable, plat ou légèrement descendant et des valeurs moyennes inférieures, en décrochage du contexte lexical environnant (Candea, 2000 ; Duez, 2003 ; Shriberg, 1994), etc. L'analyse de données issues de langues différentes laisse supposer que les hésitations vocaliques sont produites avec un contrôle articulaire plus faible que les segments phonémiques (Candea *et al.*, 2005). Cela entraîne non seulement des valeurs moyennes de  $F_0$  inférieures à celles des voyelles intralexicales, mais aussi une variabilité plus importante, reflétée dans une déviation standard plus grande. Les voyelles d'hésitation s'avèrent également sujettes à des erreurs de détection automatique de  $F_0$  plus souvent que les segments vocaliques intralexicaux. De plus, le paramètre  $F_0$  varie avec le style de parole et, comme nous avons pu le mesurer dans une étude antérieure, l'expression en L2 semble augmenter le caractère irrégulier de la hauteur des hésitations vocaliques (Vasilescu *et al.*, 2006).

### 5.1. Fréquence fondamentale des voyelles d'hésitation à travers les langues

Nous avons mesuré la hauteur des hésitations vocaliques à travers les langues. Le tableau 5 récapitule les valeurs moyennes de  $F_0$  et la déviation standard pour les voyelles d'hésitation, ainsi que les valeurs correspondantes pour les segments vocaliques intralexicaux. Pour ces derniers, la valeur moyenne de  $F_0$  est calculée comme moyenne de tous les segments par langue. Nos données confirment les observations citées ci-dessus selon lesquelles la hauteur des voyelles d'hésitation est en moyenne inférieure à celle des voyelles intralexicales. Cette tendance se vérifie pour l'anglais américain, le français et l'espagnol. À cela s'ajoute une déviation standard plus importante pour les voyelles d'hésitation, ce qui confirme les observations faites dans (Candea *et al.*, 2005)<sup>11</sup>. Enfin, nous notons également un effet de langue : la fréquence fondamentale des hésitations vocaliques atteint des valeurs moyennes plus hautes en espagnol qu'en anglais américain et en français. Nous rappelons que les données en espagnol sont légèrement différentes des données en anglais américain et en français, car il s'agit de débats parlementaires télévisés et non pas de journaux audiodiffusés, même si dans les trois langues il est question de « parole semi-préparée ».

---

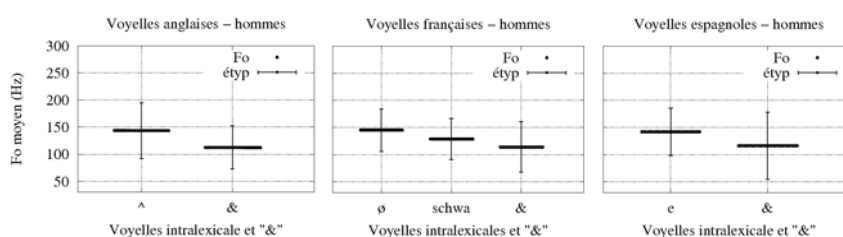
11. Afin d'étayer l'observation de Candea *et al.* (2005), nous avons évalué la distribution en fréquence des valeurs de  $F_0$  pour les hésitations vocaliques vs. les voyelles intralexicales dans les trois langues analysées. Les histogrammes (non représentés dans cet article) confirment que les valeurs de  $F_0$  des voyelles intralexicales se placent préférentiellement autour de la zone des 100 à 150 Hz, tandis que les hésitations présentent une distribution plus étalée. Cette observation est valable pour les trois langues.

	Anglais am.	Français	Espagnol
<b>Nbr. items</b>	2013	1199	1993
<b>MoyH</b>	127	125	135
<b>ETypeH</b>	72	78	57
<b>MoyV</b>	140	133	143
<b>ETypeV</b>	44	50	41

**Tableau 6.**  $F_0$  (en Hz) des voyelles d'hésitation et intralexicales en anglais américain, français et espagnol. MoyH=valeur  $F_0$  moyenne pour les voyelles d'hésitation ; ETypeH=écart type des voyelles d'hésitation ; MoyV=valeur  $F_0$  moyenne pour les voyelles intralexicales calculée comme moyenne/corpus ; ETypeV=écart type par corpus.

## 5.2. Fréquence fondamentale des voyelles d'hésitation vs. voyelles intralexicales

Afin de mieux évaluer les différences de hauteur des voyelles d'hésitation vs. intralexicales nous avons calculé la  $F_0$  moyenne pour les premières et leurs homologues phonémiques de timbre proche (figure 9). À titre d'exemple, les données utilisées en français pour cette comparaison étaient de l'ordre de grandeur suivant : 27 k segments /ə/ (/ə/+/œ/) et 1 k segments correspondant à la voyelle d'hésitation. Nos données confirment la différence de hauteur notée à propos de la voyelle d'hésitation vs. voyelles intralexicales : la  $F_0$  moyenne est plus haute pour ces dernières même si nous n'avons pas observé une différence statistiquement significative.



**Figure 9.** Moyennes de  $F_0$  en anglais américain, français et espagnol et écarts types pour les voyelles d'hésitation et intralexicales de timbres proches.

## 6. Conclusion et perspectives

Dans cette contribution nous nous sommes intéressés aux particularités acoustiques et prosodiques des hésitations vocaliques dans une perspective inter et intralangue. Nous avons examiné la question des traits communs (« universels ») déployés par les hésitations vocaliques, dont celui du timbre central et « neutre » auquel nous avons prêté une attention particulière. D'un point de vue intralangue, nous avons estimé la relation entre hésitations vocaliques et système (vocalique) de la langue.

Des hésitations vocaliques en anglais américain, français et espagnol ont été extraites automatiquement de corpus oraux de journaux radiodiffusés et de débats parlementaires. Les qualités vocaliques à travers les deux premiers formants F1 et F2, la durée et la fréquence fondamentale ( $F_0$ ) ont été mesurées et comparées à la fois du point de vue inter et intralangue. L'étude du timbre a montré que les hésitations sont réalisées de manière différente à travers les langues. Elles n'occupent pas nécessairement une position centrale dans l'espace vocalique et le timbre illustré dépend de la langue. Les voyelles d'hésitation se sont avérées proches de certains timbres de l'inventaire phonémique sans pour autant s'y confondre. En anglais américain la voyelle d'hésitation montre un timbre central plutôt ouvert entre /ʌ/ et /æ/, en français elle est centrale et un peu plus ouverte que /ə/ tandis qu'en espagnol elle occupe une position proche de la voyelle mi-fermée antérieure /e/. Afin de mieux préciser la préférence des hésitations vocaliques pour les timbres mentionnés ci-dessus, nous avons tenté de mettre en relation ce paramètre avec des études sur l'organisation des espaces vocaliques et la fréquence des phonèmes vocaliques dans les langues du monde. La question du caractère dépendant de la langue des bases articulatoires a également été considérée. Plus d'études sont nécessaires pour étayer cette piste, notamment la prise en compte d'un échantillon représentatif de langues possédant des hésitations vocaliques. Cependant nos résultats, ainsi que d'autres études consacrées au phénomène étudié semblent appuyer cette hypothèse, à savoir qu'il existe une corrélation entre l'organisation de l'espace vocalique et la présence des timbres d'hésitation dans certaines régions de cet espace.

La durée et la fréquence fondamentale ( $F_0$ ) montrent des archétypes comparables à travers les langues.

Pour ce qui est de la durée d'environ 300 ms en moyenne, elle s'avère particulièrement longue : elle est ainsi bien supérieure à la durée des segments phonémiques. La prise en compte du phénomène de centralisation des timbres avec la réduction de la durée a montré des comportements similaires pour les voyelles d'hésitation et intralexicales. D'autres analyses liées à la durée moyenne pourraient compléter les informations sur la relation entre les hésitations vocaliques et le système de la langue, comme par exemple l'effet du débit sur la réduction vocalique des voyelles d'hésitation *vs.* intralexicales. Les hésitations vocaliques se comportent-elles de la même manière que les voyelles intralexicales, *i.e.*, une

réduction vocalique, lorsque le débit est rapide ? Ou bien des voyelles intralexicales hypoarticulées (donc brèves) trahiraient un style plus spontané et donc susceptible de nécessiter des pauses de planification plus longues ?

Quant à la  $F_0$ , un décrochage entre voyelles phonémiques et voyelles d'hésitation est à noter, ces dernières montrant des valeurs moyennes inférieures. La variabilité intraclasse des hésitations vocaliques nécessite toutefois des précisions supplémentaires, notamment en ce qui concerne l'hypothèse du contrôle articulaire réduit dans la production de ces segments vocaliques.

Les perspectives de ce travail sont nombreuses. D'une part, étendre l'étude acoustique et prosodique à d'autres langues et dialectes permettrait de corroborer nos conclusions. Il serait également intéressant de considérer d'autres styles de parole pour mieux évaluer l'impact des facteurs autres que l'idiome sur les paramètres acoustiques et prosodiques caractérisant les hésitations. D'autre part, s'intéresser à la fonction dans le discours des hésitations vocaliques compte tenu des facteurs langue/dialecte et style de parole permettrait de dresser un tableau complet du phénomène. Enfin, ces résultats peuvent trouver une utilité incontestable dans le traitement automatique de la parole, car ils permettent d'affiner les modèles actuels de pauses remplies aussi bien dans le cadre de la transcription automatique que dans celui de l'identification des langues.

#### Remerciements

Ce travail a été partiellement financé par les projets suivants : MIDL (modélisation pour l'identification des langues) – projet CNRS et EC TC-STAR et CHIL – projets européens. Les auteurs tiennent à remercier Maria Candea (Paris 3 – Sorbonne Nouvelle) pour sa collaboration dans le cadre du projet MIDL.

#### 7. Bibliographie

- Adda-Decker, M., Gendrot, C., Impact of Duration on F1/F2 Formant Values of Oral Vowels : An Automatic Analysis of Large Broadcast News Corpora, *Proceedings of InterSpeech*, Lisbonne, 2005, Portugal.
- Adda-Decker, M., De la reconnaissance automatique de la parole à l'analyse linguistique de corpus oraux, *Actes des Journées d'Étude sur la Parole JEP'06*, 2006, Dinard, France.
- Adda-Decker, M., Problèmes posés par le schwa en reconnaissance et en alignement automatique de la parole, *Actes des 5<sup>e</sup> Journées d'Études Linguistiques de Nantes*, 2006, France.
- Batliner, A., Kiesling, A., Burger, S., Nöth, E., Filled pauses in spontaneous speech, *Proceedings of XIIIth International Congress of Phonetic Sciences*, 1995, Stockholm, Suède.
- Bloomfield, L., *Language*, NY, 1993.

- Boë, L.J., Perrier, P., Guérin, B., Schwarz, J.L., Maximal vowel space, *Proceedings of Eurospeech*, 1989, Paris, France.
- Boersma, P., Weenink, D., Praat, a system for doing phonetics by computer, Institute of Phonetic Sciences of the University of Amsterdam, 1999, p. 132-182.
- Boomer, D.S., Hesitation and grammatical encoding, *Language and Speech*, n°8, 1965, p. 148-158.
- Brennan, S.E., Williams, M., The feeling of another's knowing : prosody and filled pauses as cue to listeners about the metacognitive states of speakers, *Journal of Memory and Language*, 34, 1995, p. 383-398.
- Butterworth, B., Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia, *Brain and Language*, 8, 1979, p. 133-161.
- Candea, M., *Contribution à l'étude des pauses silencieuses et des phénomènes dits d'hésitation en français oral spontané*, Thèse de Doctorat de l'Université Paris 3 – Sorbonne Nouvelle, 2000.
- Candea, M., Vasilescu, I., Adda-Decker, M., Inter- and intra-language acoustic analysis of autonomous fillers, *Proceedings of Disfluency In Spontaneous Speech (DISS'05) Workshop*, 2005, Aix-en-Provence, France.
- Chafe, W.L., *The Pear Stories : Cognitive, Cultural, and Linguistic Aspects of narrative Productions*, 1980, Ablex Norwood, NJ.
- Clark, H.H., Speaking in time, *Speech Communication*, n° 36, 2002, p. 5-13.
- Clark, H.H., Fox Tree, J.J., Using uh and um in spontaneous speaking, *Cognition*, vol. 84, 2002, p. 73-111.
- van Donzel, M. E., Koopmans-van Beinum, F. J., Pausing strategies in discourse in Dutch, *Proceedings of the 4<sup>th</sup> International Conference on Speech an Language Processing*, 1996, Philadelphia, USA.
- Duez, D., Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français, *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 20, 2001, p. 31-48.
- Duez, D., Modelling aspects of reduction and assimilation in spontaneous french speech, *Actes de IEEE-ISCA Workshop on Spontaneous Speech Processing and Recognition*, 2003, Tokyo.
- Eklund, R., Shriberg, E., Crosslinguistic disfluency modelling : a comparative analysis of Swedish and American English Human-Human and Human-Machine Dialogs, *Proceedings of the 5<sup>th</sup> International Conference on Speech an Language Processing*, 1998, Sydney, Australia.
- Fougeron, C., Gendrot, C., Bürki, A. On the acoustic characteristics of the French schwa, *Proceedings of International Conference on Phonetic Sciences*, 2007, Saarbrucken, Allemagne.
- Gauvain, J.L., Lamel, L., Adda, G., The Limsi broadcast news transcription system, *Speech Communication*, n°37,vol. 1-2, 2002, p. 89-108.

- Gick, B., Wilson, I., Koch, K., Cook, C., Language specific articulatory settings : evidence from inter-utterance rest position, *Phonetica*, n° 61 vol. 4, 2004, p. 220-233.
- Goldman-Eisler, F., The determinants of the rate of speech output and their mutual relations, *Journal of Psychosomatic research*, n° 1, 1956, p. 137-143.
- Grojean, F., Deschamps, A., 1975, Analyse contrastive des variables temporelles de l'anglais et du français : vitesse de parole et variables composantes : phénomènes d'hésitation, *Phonetica*, n°31, 1975, p. 144-184.
- Lindblom, B., Phonetic universals in vowel systems, in J. J. Ohala & J. Jaeger (eds.) *Experimental Phonology*, London : Academic Press, 1986.
- Maclay, H., Osgood, C.-E., 1959, Hesitation phenomena in spontaneous English speech, *Word*, n°15, 1959, p. 19-44.
- Mahl, G.F., Disturbances in the patient's speech in psychotherapy, *Journal of Abnormal and Social Psychology*, n°42, 1956, p. 3-32.
- Marsico, E., Maddieson, I., Coupé C., Investigating the "hidden" structure of phonological systems, *Proceedings of the 30th Meeting of the BLS*, Feb 13-16, Berkeley, vol. 30, 2004, p. 256-267.
- Nemoto, R., *Hésitation vocalique autonome vs. système phonétique de la langue : étude acoustique en plusieurs langues*, Mémoire de DESS, LIMSI-CNRS, 2006.
- O'Connell, D.C., Kowal, S., The history of research on filled pauses as evidence of the written language bias in linguistics (Linell, 1982), *Journal of Psycholinguistic Research*, n°33, vol. 6, 2004, p. 459-474.
- O'Connell, S. Kowal, Dill, E.J., Dialogicality in TV news interviews, *Journal of Pragmatics*, n°36, 2004, p. 185-205.
- O'Connell, S. Kowal, *Uh and um revisited : Are they interjections for signalling delay?*, *Journal of Psycholinguistic Research*, vol. 43, n°6, 2005, p. 555-575.
- Schwarz, J.L., Boë, L.J., Vallée, N., Abry, C., The dispersion-focalization theory of vowel systems, *Journal of Phonetics*, n°25, 1997, p.255-286.
- Shriberg, E., 1994, *Preliminaries to a Theory of Speech Disfluencies*, PhD thesis, University of California at Berkeley, 1994.
- Vasilescu, I., Adda-Decker, M., Language, gender, speaking style and language proficiency as factors characterizing autonomous vocalic fillers in spontaneous speech, *Proceedings of InterSpeech*, 2006, Pittsburgh, USA.
- Vasilescu, I., Candea, M., Adda-Decker, M., Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages, *Proceedings of InterSpeech*, 2005, Lisbonne, Portugal.
- Vasilescu, I., Adda-Decker, M., A cross-language study of acoustic and prosodic characteristics of vocalic hesitations, *The Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue*, NATO Security Through Science Program, IOS Press BV, 2006, p. 140-149.

Zhao, Y., Jurafsky, D., A preliminary study of mandarin filled pauses, *Proceedings of Disfluency In Spontaneous Speech (DISS'05) Workshop*, 2005, Aix-en-Provence, France.

Watanabe, M., Den, Y., Hirose, K., Miwa, S., Minematsu, N., Factors affecting speakers' choice of fillers in Japanese presentations, *Proceedings of InterSpeech*, 2006, Pittsburgh, USA.