



HAL
open science

Que disent nos silences ? apport des données acoustiques, articulatoires et physiologiques pour l'étude des pauses silencieuses

Muriel Lalain, Thierry Legou, Camille Fauth, Fabrice Hirsch, Ivana Didirkova

► To cite this version:

Muriel Lalain, Thierry Legou, Camille Fauth, Fabrice Hirsch, Ivana Didirkova. Que disent nos silences ? apport des données acoustiques, articulatoires et physiologiques pour l'étude des pauses silencieuses. JEP - Journées d'Etude sur la Parole, Jul 2016, Paris, France. hal-01462180

HAL Id: hal-01462180

<https://hal.science/hal-01462180>

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Que disent nos silences ? Apport des données acoustiques, articulatoires et physiologiques pour l'étude des pauses silencieuses

Muriel Lalain¹, Thierry Legou¹, Camille Fauth², Fabrice Hirsch³, Ivana Didirkova³

(1) Université d'Aix-Marseille, CNRS, LPL, UMR 7309, Aix-en-Provence, France

(2) Université de Strasbourg, Institut de Phonétique, E.A. 1339 LILPa, Strasbourg, France

(3) Université Paul Valéry, Montpellier 3, CNRS, Praxiling, UMR 5267, Montpellier, France

muriel.lalain@lpl-aix.fr

RESUME

Si la rhétorique s'est intéressée très tôt à la pause, il a fallu attendre le XX^{ème} siècle pour que d'autres disciplines – la psycholinguistique, le traitement automatique des langues, la phonétique – accordent à ces moments de silence l'intérêt qu'ils méritent. Il a ainsi été montré que ces ruptures dans le signal acoustique, loin de signifier une absence d'activité, constituaient en réalité le lieu d'une activité physiologique (la respiration) et/ou cognitive (planification du discours) qui participent tout autant au message que la parole elle-même.

Dans cette étude pilote, nous proposons des observations et des pistes de réflexions à partir de l'analyse des pauses silencieuses dans un corpus de parole lue et semi dirigée. Nous mettons notamment en évidence l'apport de l'analyse conjointe de données acoustiques, articulatoires (EMA) et physiologiques (respiratoires) pour l'identification, parmi les pauses silencieuses, des pauses respiratoires, syntaxiques et d'hésitation.

ABSTRACT

What do our silences say? Contribution of acoustic, articulatory and physiological data to the study on silent pauses.

While rhetoric has been interested in pauses since a long time, researches on those brief moments of silence in other scientific fields, such as psycholinguistics, natural language processing or phonetics were rather rare before the 20th century. Thus, it has been shown that ruptures in acoustic signal are not a sign of a lack of activity; they are in fact the place of a physiological (respiration) and / or cognitive (discourse planning) activity. Both are as important for the message as the speech activity.

In this pilot study, observations and reflections are made, based on analysis of silent pauses in semi-directed and read speech. Our research shows the importance of considering the acoustic, articulatory (EMA) and physiological (Biopac) data together, in order to identify respiratory, syntactic and hesitation pauses within the brief silences.

MOTS-CLES : pauses ; mouvements articulatoires ; EMA ; Biopac ; respiration ; déglutition

KEYWORDS: pauses ; articulatory movements ; EMA ; Biopac ; respiration ; deglutition

1 Introduction

La pause fait partie intégrante des tours de parole. A ce titre, elle a toujours suscité un certain intérêt depuis l'Antiquité. Depuis le milieu du XX^{ème} siècle, linguistes et psycholinguistes ont également mené une réflexion sur ces événements présents dans la parole. Différents travaux ont ainsi rendu compte de la quantité des pauses dans le discours, leur durée, leur nature et leurs fonctions. Sur le plan phonétique, deux types de pauses peuvent être relevés : les pauses pleines et les pauses vides. Les premières correspondent à des allongements de sons, à la réalisation d'un *schwa* (Maclay et Osgood, 1959) ou à l'utilisation d'un autre type de filler (« hum » ou « bin »). Les pauses vides, elles, peuvent être définies comme une interruption du flux de parole se répercutant sur le signal acoustique par une amplitude nulle ou non-significative (Duez, 2003).

D'après Goldman-Eisler (1968) les pauses constituent près de 50% du temps de parole lors d'une description d'image. Ce bref moment de silence est ainsi le signe d'une activité cognitive importante puisque la pause permet de reprendre sa respiration (pauses physiologiques), de planifier le contenu de son message et structurer son énoncé (pauses syntaxiques), et de procéder à une recherche et sélection lexicale (pauses d'hésitation). Grosjean & Deschamps (1972) ont pour leur part montré que les pauses d'hésitation étaient moins fréquentes que les pauses syntaxiques, mais ce résultat est à relativiser en fonction de la situation de parole (lecture *vs* description) et du locuteur lui-même, selon sa plus ou moins grande habileté en lecture par exemple (Lalain et al, 2014). La durée des pauses vides a également suscité un certain intérêt. Grosjean & Deschamps (1975) ont montré qu'elles durent en moyenne 520 ms en parole spontanée et 1320 ms lorsqu'il s'agit de décrire une image. Ces résultats, tout comme ceux de Duez (2001) ou de Goldman et al (2010) montrent que la durée de la pause dépend du style de parole (lecture, parole conversationnelle,...). Cette dernière étude a également révélé que l'étendue de la pause était fonction de la présence ou non d'une prise de respiration et/ou d'une déglutition de salive, tandis que l'étude de Hirsch *et al.* (2015) a montré que les caractéristiques temporelles des pauses peuvent également dépendre de la charge émotionnelle et de la thématique abordée. Les durées des pauses peuvent également dépendre de l'activité motrice qui s'y déroule : l'étude articulatoire et perceptive d'Abry *et al.* (1996) a ainsi montré que des gestes anticipatoires d'arrondissement des lèvres ou d'abaissement de la mandibule pouvaient être présents dans les pauses, mais ce phénomène serait partiellement lié au locuteur et à la durée de la pause.

Ainsi, les travaux sur la pause ont permis de mieux connaître leur nature (syntaxique, respiratoire, recherche lexicale, focalisation...), leur distribution ou encore leur durée, et ce, en fonction de différents styles de parole. Il est ainsi admis en règle générale deux grands types de pauses, et ce dans tous les styles de parole : les pauses syntaxiques et les pauses d'hésitation dont on sait qu'elles se distinguent à la fois par leur fréquence d'apparition, leur durée et leur localisation syntaxique. Cependant, les interruptions sonores d'ordre syntaxique recouvrent également les pauses dont la fonction peut être physiologique, comme les pauses respiratoires ou celles dévolues à la déglutition.

Notre objectif, avec cette étude pilote, est de prolonger les recherches menées sur la pause en élaborant un protocole expérimental fiable, permettant d'utiliser des données à la fois acoustiques, articulatoires et physiologiques pour l'identification objective des différents types de pauses suscités. En d'autres termes, ce protocole visera à mettre en évidence d'éventuelles différences acoustiques, articulatoires et / ou physiologiques entre les différentes catégories de pauses décrites *supra*.

2 Méthodologie

2.1 Recueil des données

Pour cette étude pilote, nous avons constitué un corpus de parole auprès d'une locutrice volontaire au Laboratoire Parole et Langage à Aix-en-Provence. Celle-ci ne présentait ni trouble de l'audition ni trouble de la parole ou du langage. Nous avons recueilli des données acoustiques, articulatoires et physiologiques afin de pouvoir utiliser ces différents indices dans la caractérisation des pauses.

Données Audio, Articulatoires (EMA) et Physiologiques (Biopac)

Les enregistrements audio et articulatoires ont été réalisés avec un articulographe AG501 (5D) à une fréquence de 250 Hz et avec une précision meilleure que le mm sur les trois axes (X, Y, Z). Les huit capteurs utilisés ont été plongés dans du latex avant d'être imprégnés de colle et positionnés comme suit : un capteur sur la mastoïde gauche (1), un sur la lèvre supérieure (2), un sur la lèvre inférieure (3), un à l'inter-incisives supérieure (4), un à l'inter-incisive inférieure (5), deux sur la langue (dos 6 et apex 7), un au niveau du larynx (8). Le capteur placé au niveau du larynx a été collé sur la peau après repérage de la position haute de la partie saillante du larynx lors d'une déglutition. Les capteurs 1 et 4 sont des capteurs de référence permettant la correction des mouvements de tête, les autres capteurs fournissent des indications sur la position et les mouvements des articulateurs. Une ceinture thoracique du type SS5LP connectée à la station d'acquisition MP35 de la société Biopac a été utilisée pour la visualisation de l'activité respiratoire par la variation du périmètre de la cage thoracique.

Synchronisation des enregistrements

L'AG501 intègre un module de synchronisation (SyBox) qui assure d'emblée la synchronisation entre l'enregistrement audio du locuteur et celui des signaux des capteurs de positions. Pour synchroniser l'articulographe avec le Biopac, nous avons développé un module à base d'un microcontrôleur (Arduino). Lorsqu'il reçoit un signal de la SyBox, ce module génère un bip audio (500Hz pendant 500ms) et un signal de déclenchement à destination du Biopac. Le Biopac est configuré pour « attendre » le signal du microcontrôleur avant de commencer l'acquisition. Les données articulatoires, audio et respiratoires sont recueillies à différentes fréquences mais sont ainsi synchronisées.

Corpus

Le sujet devait accomplir 7 tâches (T1 à T7). Les consignes étaient présentées sur un écran d'ordinateur (ppt). Les T1 et T2 consistaient à lire à haute voix et à un rythme usuel *La chèvre de Monsieur Seguin* et un texte intitulé *Pascal*. Ces deux tâches de lecture ont été réitérées en fin d'expérimentation (T6 et T7) dans le but de vérifier un éventuel effet d'habituation au dispositif EMA. Deux tâches de description d'images suivies d'une discussion informelle avec l'un des expérimentateurs ont également été proposées. Il s'agissait de décrire La Une de Charlie Hebdo qui a suivi les attentats du 13 novembre (T3) et des photos de Robert Doisneau (T4) représentant des scènes quotidiennes à Paris. La discussion initiée par l'expérimentateur était centrée sur les ambiances terreur/quiétude représentées par ces différents supports. L'objectif des T3 et T4 était d'obtenir une parole semi contrôlée (description) et conversationnelle (discussion) à l'inverse des T1 et T2 où la production de parole est contrôlée.

2.2 Traitement des données

Les données recueillies au cours de cette étude pilote constituent un corpus d'une durée d'environ 20 mn. Les tâches de lecture, description/discussion et déglutition permettent l'observation de divers phénomènes parmi lesquels la fréquence d'occurrence des pauses, le type de pause et leur durée.

Identification des pauses et de leur nature

L'ensemble du corpus a été transcrit orthographiquement (sous Praat), puis segmenté et annoté de manière semi-automatique (Easy Align). Seules les tiers syllabes et ortho ont été conservées du traitement fait à l'aide d'Easy Align. Les pauses ont été annotées dans la tier Syllabes ; une tier scope a été ajoutée pour l'annotation de segments comprenant la pause ainsi que les syllabes pré et post-pausales.

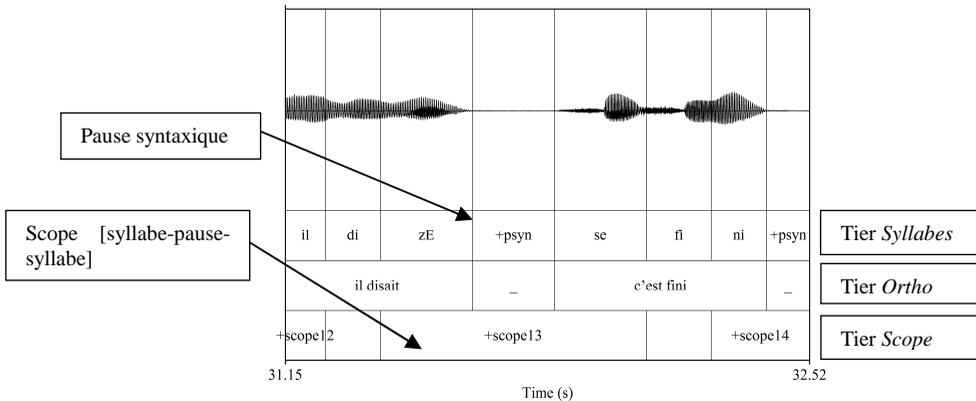


Figure 1 : Etiquetage du signal acoustique avec l'annotation d'une pause syntaxique et d'un scope.

Cette analyse concerne uniquement les pauses vides présentes dans les tâches 1, 2, 3, 4, 6 et 7. Ces pauses ont été identifiées à l'aide d'indices perceptivo-visuels : elles correspondent à un silence perceptible accompagné d'une rupture d'activité acoustique significative visible sur le signal de parole (cf. Fig. 1). Aucun seuil de durée n'a été appliqué, contrairement à ce qui est habituellement opéré : l'un des objectifs de cette étude pilote était de vérifier l'apport des données acoustiques articulatoires et physiologiques dans l'identification et la caractérisation des pauses. Seuls ont été éliminés les silences correspondant aux tenues des occlusives sourdes et aux occlusives glottales. La durée de la tenue des occlusives sourdes en position initiale ou post-pausale a été déterminée à partir de la durée moyenne de la tenue de l'ensemble des occlusives sourdes en contexte VCV. Les pauses vides syntaxiques ont été prédites à partir de la fonction POS de Marsatag (pour une description plus complète de cet outil d'enrichissement de données textuelles et de transcriptions de l'oral, voir Rauzy et al., 2014) ainsi que de la ponctuation pour les textes lus. Nous avons ainsi pu identifier, parmi les pauses vides, celles dont la fonction était syntaxique (+psyn) et celles qui relevaient d'autocorrections (+pac) et de cas d'erreurs de décodage (+pdec) en lecture ou d'hésitations avec (+prhes) ou sans (+phes) activité respiratoire en description/discussion. L'examen des signaux de respiration correspondant à chacune des +psyn a permis d'identifier de manière objective les pauses silencieuses respiratoires recodées +psryn. La figure 2 représente les différents graphes à partir desquels nous avons déterminé la nature des différentes pauses et basé nos observations. Pour

chaque pause, nous avons généré une figure pour l'analyse des mouvements articutoires selon l'axe Z (élévation-abaissement) et pour l'analyse du signal de respiration. De haut en bas, les graphes 1 à 4 représentent les positions des articulateurs mâchoire/apex/dos /larynx, en Z ; Sur l'axe vertical est représentée l'amplitude du mouvement, le point 0 correspondant à la position de l'articulateur en début de pause. Le graphe 5 rapporte le signal délivré par la ceinture permettant d'observer l'augmentation du périmètre thoracique, le graphe 6 représente le signal audio enregistré. Tous les graphes correspondent à la même durée d'observation, laquelle correspond à la durée des pauses, représentées en secondes sur l'axe commun à tous les graphes au bas de la figure.

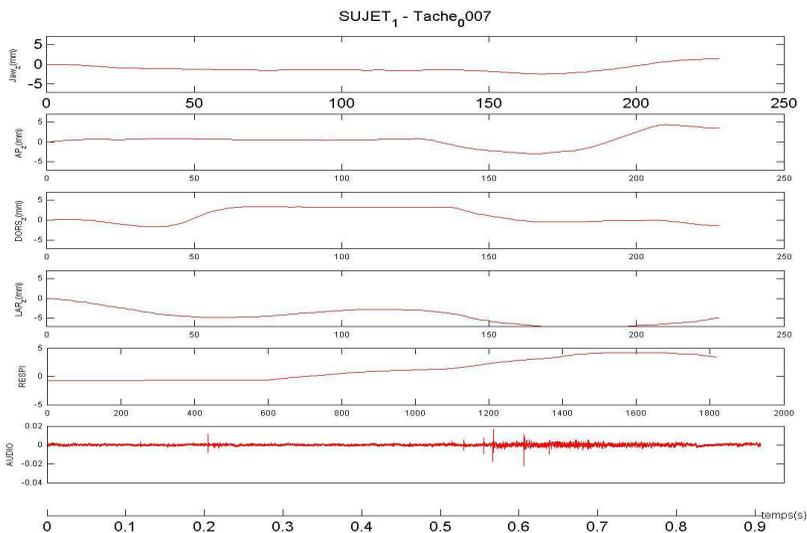


Figure 2 : Graphes utilisés pour la caractérisation des pauses et l'identification de patterns articutoires et physiologiques ; Ici une pause syntaxique respiratoire (augmentation du périmètre thoracique) précédant le segment [ʒ] (élévation de la mâchoire et de l'apex)

L'utilisation des données acoustiques, syntaxiques et physiologiques nous a ainsi permis d'identifier, parmi les pauses vides initialement relevées, des sous catégories selon les tâches : les pauses vides syntaxiques (+psyn), les pauses vides correspondant à des autocorrections (+pac) ou des erreurs de décodage (+pdec) les pauses vides d'hésitation avec (+prhes) ou sans (+phes) activité respiratoire et les pauses vides syntaxiques respiratoires (+prsyn).

Fréquence d'apparition et durée des pauses selon leur nature et par tâche

Après avoir identifié les différents types de pauses, nous avons examiné leur fréquence d'apparition, à partir de la tier syll qui comprend le codage +p*. Les différentes natures des tâches accomplies par le sujet (lecture vs description/discussion) ont eu pour conséquence la production de parole de durées différentes : les deux tâches de lecture et la tâche de description ne comprennent pas le même nombre de syllabes produites (T1 = 277 syllabes vs T2 = 110 syllabes vs T3= 240). Afin de pouvoir comparer le nombre de pauses vides entre ces différents exercices de parole, nous avons calculé leur probabilité d'apparition pour chaque tâche : nbpause/(nbsyll+nbpause).

3 Observations préliminaires

3.1 Fréquences d'occurrences des pauses silencieuses

La figure 3 donne les fréquences d'occurrence des différents types de pauses relevés dans chaque tâche. Les pauses silencieuses syntaxiques et les pauses silencieuses syntaxiques respiratoires sont relevés dans les tâches 1 à 7. Les pauses d'autocorrection et de décodage seulement en T6 et T7, et les pauses d'hésitation avec (prhes) et sans (phes) activité respiratoire seulement en T3 et T4.

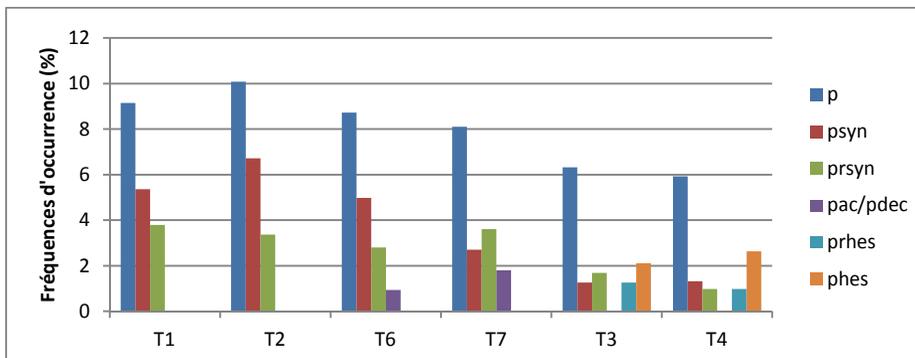


Figure 3 : Fréquences d'apparition (en %) des différents types de pause par tâche

La fréquence d'apparition des pauses syntaxiques est supérieure à celle des pauses respiratoires dans les tâches de lecture alors que cette tendance est inversée en T3 et T4. Ceci peut s'expliquer par le fait qu'en lecture, les pauses syntaxiques sont fortement contraintes, notamment en partie par la ponctuation. Concernant les comparaisons T1 vs T6 et T2 vs T7 rien n'indique une habitude au dispositif EMA, d'autant qu'il n'y a pas de différence en ce qui concerne le débit et la vitesse articulaire.

3.2 Durées des pauses

L'intervalle +ps a été utilisé pour calculer la durée des pauses silencieuses dans chaque tâche. Les durées moyennes des pauses sont données pour chaque type de pause pour les 6 tâches, en secondes.

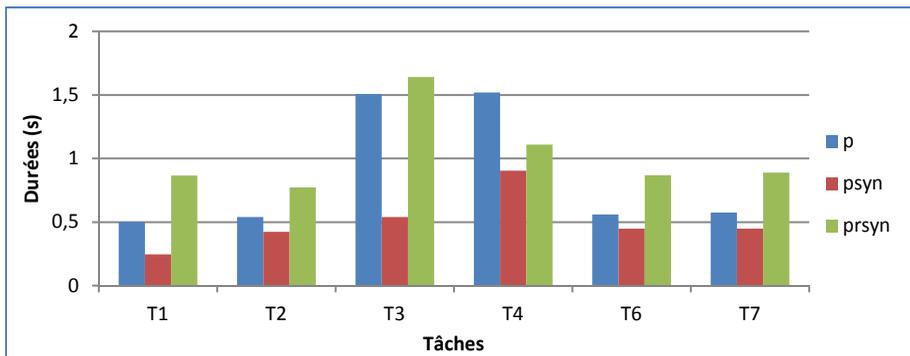


Figure 5 : Durée moyenne (s) des pauses par type de pause et par tâche

On retrouve une observation connue dans la littérature : les pauses sont plus longues en T3 et T4 (tâches semi contrôlées) que dans les autres tâches. Cette tâche de génération de parole implique en effet une charge cognitive plus intense que l'activité de lecture, ce qui se traduit davantage par une durée plus importante qu'une fréquence d'apparition plus élevée. Cet examen de la durée montre également une différence de durée des pauses pour chacune des cinq tâches, entre les pauses syntaxiques et les pauses syntaxiques respiratoires. Ces dernières sont nettement plus longues.

3.3 Patterns physiologiques et articulatoires

A partir des distinctions faites concernant les types de pauses, nous nous sommes intéressés aux articulateurs, en particulier la mandibule, la langue (apex et dos) ainsi que le larynx, sur l'axe Z (plan vertical), et ce afin de repérer des patterns articulatoires récurrents correspondants aux pauses silencieuses respiratoires et aux pauses silencieuses syntaxiques. Cette étape est un préalable à une réflexion concernant le développement d'une méthode d'analyse permettant l'objectivation de ces possibles patterns.

Le capteur 8 positionné sur la peau du cou au niveau du cartilage thyroïde lorsque celui-ci est dans sa position haute en déglutition permet de repérer l'ascension du larynx pendant le deuxième temps de la déglutition. Cette montée du larynx s'accompagne d'une élévation de l'apex et du dos de la langue qui correspondent à l'appui lingual antéro postérieur observable en fin de temps buccal et qui permet la propulsion de la salive vers le pharynx. Ce pattern articulatoire semble récurrent en lecture et en description/narration pour signer les déglutitions physiologiques, lesquelles ont lieu au cours des pauses syntaxiques, en l'absence de respiration (puisque la déglutition s'accompagne d'une inhibition respiratoire pendant les temps oral et pharyngé). Cependant, cette mesure à l'aide du capteur 8 demande à être validée par une étude spécifique visant à préciser la fiabilité de ce capteur et de son positionnement.

A partir des intervalles +scope correspondant à la dernière syllabe précédant la pause et la première suivant la pause (syll +ps syll), nous avons récupéré les mouvements articulatoires pré et post pausaux, afin de pouvoir repérer d'éventuels phénomènes d'anticipation articulatoire pouvant être observés pendant les pauses. Nous avons ainsi observé des patterns articulatoires associés aux pauses silencieuses respiratoires qui semblent refléter les gestes d'anticipation du segment suivant. On peut par exemple relever en fin de pause le mouvement d'élévation de la mandibule et de l'apex que l'on peut attribuer à la préparation articulatoire de la consonne fricative [3] (cf Figure 2)

4 Conclusion

L'objectif de cette étude pilote était de montrer la pertinence de l'utilisation conjointe de données acoustiques, articulatoires et physiologiques pour l'identification et la caractérisation des pauses silencieuses dans différents styles de parole. Les observations préliminaires présentées ici ont mis en évidence l'intérêt de l'apport des données physiologiques dans l'identification des pauses silencieuses respiratoires. En effet, on a pu observer que toutes les pauses syntaxiques n'ont pas un rôle physiologique respiratoire et que cette distinction permet de mettre en évidence des caractéristiques temporelles jusqu'ici ignorées : les pauses respiratoires sont en effet plus longues que les pauses syntaxiques. Les données articulatoires semblent pouvoir également contribuer à l'identification d'indices robustes pour caractériser la déglutition (élévation de l'apex, du dos et du larynx). De plus, ces données pourraient contribuer à distinguer les pauses silencieuses syntaxiques et d'hésitation, par des patterns articulatoires différents (anticipation ou recherche articulatoire).

Ces premières observations nous conduisent donc à considérer notre protocole expérimental comme valide, même si quelques améliorations sont à envisager avant le recueil de données auprès d'une plus large population. Pour les tâches de lecture et de discussion, une pré-tâche permettrait d'éviter les hésitations en lecture et les difficultés dans la discussion sur les images. Pour aller plus loin dans l'identification des différents types de pauses (notamment syntaxiques et d'hésitation), une utilisation d'autres fonctions de l'aligneur syntaxique Marsatag pourra s'avérer utile. Le positionnement du capteur 8 au niveau du larynx, devrait faire l'objet d'une étude particulière visant à valider son degré de précision ; cela étant, les observations faites sur ce premier jeu de données sont encourageantes. Enfin, un travail de recherche et développement devrait également être centré sur l'obtention de volumes d'air inspirés et expirés corrélés aux variations de volume de la cage thoracique enregistrés par le Biopac. Ce travail pourrait être conduit avec des mesures associées de Biopac et d'EVA.

Remerciements

Cette recherche a été en partie financée par un IdEx Attractivité de l'Université de Strasbourg « Arythmique » attribué à Camille Fauth, ainsi que par le projet SYNABE (Défi 2016 : Instrumentation aux limites), porté par Fabrice Hirsch.

Références

- ABRY C., CATHIARD M.A., EL ABED R., LALLOUACHE M.T., LEROY M.C., PERRIER P., POVEDA F., SAVARIAUX C. (1996). Silent speech production: anticipatory behaviour for 2 out of the 3 main vowel gestures/features while pausing. *First ETRW on speech production modeling*, Autrans, France Mai 1996, p 101-104.
- DUEZ D., (2001). Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, vol. 20, p 31-48.
- DUEZ D. (2003). Le pouvoir du silence et le silence du pouvoir : comment interpréter le discours politique, *MediaMorphoses*, 8, p. 77-82.
- GOLDMAN-EISLER F. (1968). *Psycholinguistics. Experiments in spontaneous speech*, New York, Academic Press, 169 p.
- GOLDMAN J.-P., FRANÇOIS T., ROEKHAUT S., SIMON A.-C. (2010). Étude statistique de la durée pausale dans différents styles de parole, *Actes des XXVIIIèmes Journées d'Etude sur la Parole, Mons, 25-28 mai*, p.161-164.
- GROSJEAN F., DESCHAMPS A., (1972). Analyse des variables temporelles du français spontané, *Phonetica*, 26, 3, 129-156.
- GROSJEAN François & DESCHAMPS Alain, 1975, « Analyse contrastive des variables temporelles de l'anglais et du français », *Phonetica*, 31, 144-184.
- HIRSCH F., PEREA F., STEUCKARDT A., VERINE B. (2015). La rédemption est dans la pause. Émotion et prosodie dans l'interview de DSK au vingt heures du 18 septembre 2011, in *Comment les médias parlent des émotions. L'affaire Nafissatou Diallo contre Dominique Strauss-Kahn*, in A. Rabatel, M. Monte et M. Das Graças Soares Rodrigues (Eds), Lambert Lucas, p. 177-194.
- LALAIN M., MENDONCA-ALVES L., ESPESSER R., GHIO A., DE LOOZE C., REIS C., (2014). Prosody and reading : temporal and melodic particularities in dyslexic child in reading and narration. *Rev. Laryngol. Otol. Rhinol*, 135, 1, p. 71-82
- MACLAY H., OSGOOD C. (1959). Hesitation Phenomena in Spontaneous English Speech, *Word*, 15, p. 19-44.