

Préconisations pour la conservation et l'exploitation de données collectées pour l'étude des troubles de la voix et de la parole

G. Pouchoulin^{1,2}, A. Ghio², L. Crevier³, C. Fougeron³, C. Fredouille¹
S. Pinto², A. Giovanni², D. Robert², A. Simon³, B. Teston², F. Viallet²

¹Université d'Avignon, CERI/LIA, Avignon (France)

²Laboratoire Parole et Langage, CNRS, Université Aix-Marseille, Aix en Provence (France)

³Laboratoire de Phonétique et Phonologie, UMR 7018 CNRS-Paris3/Sorbonne Nouvelle, Paris (France)

alain.ghio@lpl-aix.fr, corinne.fredouille@univ-avignon.fr, cecile.fougeron@univ-paris3.fr

Dans le domaine de la phonétique clinique, les études nécessitent d'une organisation structurée de données variées et nombreuses pour pouvoir généraliser des régularités et des déviations empiriques observées sur une population particulière. Cela est d'autant plus vrai dans le cas de patients dysarthriques où une très forte variabilité inter-individuelle est constatée (progressions différentes de la maladie, mécanismes compensatoires individuels, symptomatologies propres à chaque patient...). Or, on constate que la plupart des études dans ce domaine reposent sur des cohortes de sujets souvent trop hétérogènes pour pouvoir représenter une population clinique particulière. Notons aussi que la rareté de certaines pathologies ne favorise pas la disponibilité des patients à participer à des tâches d'élocution en milieu clinique. De plus, la dissociation des données médicales (état clinique du patient, contexte expérimental...) des données sonores rend ces deux sources d'informations inexploitable pour une utilisation ultérieure. Pour ces raisons, la conservation pérenne et la mutualisation des ressources existantes apparaissent comme primordiales et imposent de définir clairement la normalisation et la structuration des données sur les locuteurs et leurs productions langagières.

Le travail présenté ici s'inscrit en première partie du projet DesPho-APaDy soutenu par l'ANR dont l'objectif est de décrire les caractéristiques phonéto-acoustiques de la parole dysarthrique. Il s'agit de la mise en oeuvre d'une base de données accessible via une plate-forme web dotée d'un accès sécurisé sur des données sonores, physiologiques et cliniques. Pour cela, les ressources constituant le corpus de parole « pathologique » sont constituées de :

- ◊ 1530 patients dysphoniques enregistrés avec l'appareillage EVA [1] et sur des cassettes D.A.T. (service ORL du CHU de la Timone à Marseille [MTO]);
- ◊ 990 patients dysarthriques et 160 sujets contrôles enregistrés avec l'appareillage EVA (service de neurologie du CH du Pays d'Aix [AHN]);
- ◊ 25 personnes atteintes du syndrome cérébelleux (service du Dr Claude Chevré-Muller [CCM] et [MTO]), 45 de la maladie de Parkinson ([CCM] et [AHN]), 39 de la sclérose latérale amyotrophique (Hôpital Européen Georges Pompidou) et 8 de la maladie lysosomale (Hôpital de la Salpêtrière).

Conceptualisée sous un modèle relationnel, la base de données est composée d'une cinquantaine de tables décrivant les locuteurs avec des informations civiles (date et lieu de naissance, lieu de résidence, ...), socio-linguistiques (langue maternelle, professions, ...), médicales (prescriptions thérapeutiques, recueil des symptômes, diagnostics établis, ...) et sur les sessions d'enregistrements. Contrairement aux bases de données orales telles que celles disponibles au CRDO [2], la parole « pathologique » nécessite la collecte et la mémorisation précise d'informations sur les locuteurs et sur le contexte d'enregistrement. Par conséquent, il est recommandé de recueillir le plus de renseignements sur les aspects socio-linguistiques, médicaux, symptomatiques et contextuels. Ces préconisations essentielles ne dispensent pas de l'obligation légale d'une déclaration à la Commission Nationale de l'Informatique et des Libertés [3] pour le recueil, la saisie des informations cliniques et l'analyse statistique des données recueillies au cours de la recherche. Soulignons aussi que l'établissement de consentements éclairés auprès des locuteurs ainsi que l'anonymisation des données représentent des devoirs juridiques à ne pas négliger.

[1] A. Giovanni, N. Estublier, D. Robert, B. Teston, M. Zanaret, M. Cannoni, 1995. Evaluation vocale objective des dysphonies par la mesure simultanée des paramètres acoustiques et aérodynamiques à l'aide de l'appareillage EVA. *Ann. Otolaryngol. Chir. Cervicofac.*, 112, p. 85-90.

[2] B. Bel, P. Blache Le Centre de Ressources pour la Description de l'Oral (CRDO), 2006. *Travaux interdisciplinaires du LPL d'Aix-en-Provence. (TIPA)*, vol. 25, p. 13-18.

[3] CNIL, 2006. *Méthodologie de référence pour les traitements de données personnelles opérés dans le cadre des recherches biomédicales.* MR-001, www.cnil.fr.