Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

iHist et iScatter, outils en ligne d'exploration interactive de données : application aux valeurs aberrantes de f0 et de formants

Nicolas Audibert[1]
(1) Laboratoire de Phonétique et Phonologie (CNRS & Sorbonne Nouvelle),
4 rue des Irlandais, 75005 Paris, France
nicolas.audibert@sorbonne-nouvelle.fr

RESUME

Les mesures aberrantes d'un point de vue statistique (outliers) doivent être traitées avec précaution, ce qui peut être compliqué en pratique lorsque la quantité de données devient importante. Afin de faciliter l'inspection des valeurs situées à la marge des distributions, nous proposons deux outils développés avec R/Shiny, disponibles sous forme d'applications en ligne utilisables par des non-spécialistes et distribués gratuitement sous licence GPL. Ces applications permettent de paramétrer la visualisation et d'explorer de façon interactive des distributions via des histogrammes, et les relations entre variables quantitatives via des nuages de points. Deux cas d'utilisation appliqués à des données de parole sont présentés pour illustrer les principales fonctionnalités de ces outils, à partir de mesures acoustiques extraites par Praat : l'ajustement des valeurs limites pour la détection automatique de la fréquence fondamentale, et l'identification de valeurs erronées de formants.

ABSTRACT

Online tools for interactive exploration of speech data: application to outliers in f0 and formants values.
Statistical outliers need to be handled with care, which can be difficult with large datasets. To facilitate the inspection of values at the margins of distributions, we offer two tools developed with R/Shiny, available as online applications for use by non-specialists and freely distributed under the GPL license. These applications let the user set visualization parameters and interactively explore distributions via histograms, and relationships between quantitative variables via scatterplots. Two use cases applied to speech data are presented to illustrate the main functionalities of these tools, based on acoustic measurements extracted with Praat: adjustment of limit values applicable to automatic fundamental frequency detection, and identification of erroneous formant values.

MOTS-CLES : données, valeurs aberrantes, outils en ligne, exploration interactive, f0, formants.
KEYWORDS : data, outliers, online tools, interactive exploration, f0, formants.

1 Introduction

L'analyse de données de parole nécessite fréquemment un « nettoyage » des données afin d'éliminer les mesures erronées, tout particulièrement dans le cas de mesures acoustiques (semi-)automatisées, et ce d'autant plus que les ensembles de données traités sont de grande taille. Plusieurs approches

Actes des 35èmes Journées d'Études sur la Parole, pages 598–607.

ont pu être proposées pour cela, la plus courante étant fondée sur la définition de seuils au-delà desquels les valeurs sont considérées comme erronées et donc éliminées. En psycholinguistique, ces seuils sont généralement définis à partir de critères statistiques, le plus souvent en éliminant les valeurs situées à plus de deux écarts-types de la moyenne sous l'hypothèse d'une distribution normale (ce qui revient à éliminer les 2,3% des valeurs les plus petites et les 2,3% les plus grandes). Notons toutefois que de tels seuils statistiques sont fortement dépendants de la distribution des données : dans le cas de mesures de durée pour lesquelles il est courant d'observer des distributions asymétriques, l'utilisation de ce critère qui implique un postulat de normalité conduirait ainsi à considérer à tort des valeurs élevées comme déviantes et des valeurs faibles comme non-déviantes.

Parmi les travaux récents qui se sont penchés sur la question des stratégies applicables face aux données extrêmes, on peut mentionner ceux de Nicklin & Plonski (2020) qui se concentrent plus spécifiquement sur le cas des tâches de lecture en acquisition L2 et comparent deux stratégies statistiques de résolution, l'élagage des valeurs extrêmes ou leur remplacement par des valeurs limites (*winsorizing*), concluant que le choix de l'une ou l'autre de ces stratégies a peu d'incidence sur les résultats. Quant à eux, Osborne & Overbay (2019) soulignent que si l'élimination des valeurs aberrantes permet d'estimer les effets étudiés avec à la fois plus de précision et de robustesse, toutes les valeurs extrêmes ne doivent pas être considérées comme aberrantes et inversement. Si une telle approche se justifie lorsque l'objectif est de caractériser une tendance majoritaire dans un groupe ou une condition expérimentale, les valeurs « déviantes » d'un point de vue statistique peuvent être tout aussi informatives à condition qu'elles ne relèvent pas simplement d'erreurs de liées au recueil des données ou aux mesures effectuées. Par ailleurs les erreurs ou biais de mesure ne sont pas toujours simples à caractériser à partir de la définition de seuils, notamment lorsque ces seuils doivent être établis sur la base de valeurs normatives. En effet dans le cadre de données de parole, de telles normes lorsqu'elles existent et ne se limitent pas à des valeurs moyennes sont souvent difficilement transposables à d'autres conditions de production que celles utilisées pour les établir.

S'il semble raisonnable de recommander une inspection systématique de l'intégralité des données pour identifier les cas qui constituent des erreurs, et le cas échéant distinguer parmi ces erreurs celles susceptibles de faire l'objet d'une correction des cas à éliminer de l'analyse, une telle approche est d'autant plus difficilement applicable que le volume de données traité est important. Or, avec l'évolution des moyens techniques dans les dernières décennies et l'accessibilité croissante des méthodes automatiques en sciences de la parole, la taille des corpus analysés tend à augmenter, notamment lorsque ces corpus consistent en des enregistrements acoustiques. Si cette augmentation de la taille des données a le mérite de permettre la prise en considération dans les travaux en sciences de la parole de phénomènes propres à la parole continue (Liberman, 2019) elle peut aussi accroitre le risque de ne considérer les données de parole que comme des ensembles de valeurs numériques en atténuant voire en perdant le lien direct avec l'interprétation phonétique de ces données.

Lorsque le volume de données devient conséquent et rend impossible l'inspection de l'intégralité des données, on peut recommander d'effectuer ces vérifications sur un échantillon aléatoire, et de façon plus systématique sur certaines plages de valeurs définies à partir de l'observation de la distribution de l'ensemble des données pour un groupe de locuteurs ou une condition particulière. Pour certaines analyses, les données à explorer en priorité peuvent être définies par la combinaison des valeurs prises par plusieurs variables plutôt qu'à partir de la distribution d'une variable unique, afin d'identifier les observations qui ne suivent pas la tendance majoritaire.

Dans de tels cas, un outil tel que le logiciel R combiné au regroupement de paquets *tidyverse* (Wickham et al., 2019) - ou des bibliothèques Python telles que *pandas* ou *Polars* - peut s'avérer précieux pour déterminer quels sous-ensembles de données nécessitent un examen plus approfondi.

Cependant et bien que l'initiation à l'utilisation de ces outils soit de plus en plus largement intégrée dans les formations initiales en linguistique et considérée comme un prérequis pour des formations plus avancées en statistiques, on peut constater que dans les faits leur utilisation reste souvent restreinte à des cas d'utilisation spécifique (le plus souvent la réalisation de tests statistiques), l'exploration des données s'appuyant sur des tableurs moins adaptés au jeux de données de taille conséquente. C'est ce constat qui a en partie motivé le développement des outils présentés dans cet article, initialement mis en place pour l'enseignement des statistiques à un public d'étudiants en licence et master en sciences du langage puis étendus ensuite afin de pouvoir être utiles à la communauté de recherche en sciences de la parole. Les deux applications présentées, qui permettent aux utilisateurs de téléverser leur propre jeu de données afin de l'explorer via des visualisations paramétrables, ont été développées à l'aide du logiciel R (R Core team, 2024) et du paquet shiny (Chang et al., 2024), et peuvent être utilisées directement en ligne ou localement. Nous présentons ici certaines fonctionnalités de ces outils, en les illustrant à travers des applications potentielles à des mesures acoustiques parmi celles les plus fréquemment utilisées dans les études phonétiques, la fréquence fondamentale f0 et les formants.

2 Limites des mesures automatiques de f0 et de formants

2.1 Fréquence fondamentale (f0)

La mesure de la fréquence fondamentale est réputée fiable sur des enregistrements acoustiques de parole normophonique réalisés dans des conditions permettant d'obtenir un rapport signal/bruit élevé. Toutefois, les résultats en condition non-bruitée de l'évaluation réalisée par Jouvet & Laprie (2017) indiquent que même dans ces conditions supposées idéales, l'ensemble des algorithmes évalués avec les paramètres par défaut commettent des erreurs concernant la décision de voisement ou divergent de plus de 20% de la f0 de référence (taux global d'erreur FFE compris entre 2,5% et 13% des trames analysées selon les locuteurs et algorithmes). Notons qu'une étude récente (Vaysse et al., 2022) visant à comparer les performances d'algorithmes de f0 sur des voix pathologiques suggère de combiner deux algorithmes évalués comme optimaux pour respectivement la décision de voisement et l'estimation de f0 afin d'améliorer les performances globales de la détection de f0, ce qui pourrait conduire à un taux d'erreur plus faible également sur la parole normophonique.

La majorité de ces algorithmes de détection de f0 étant paramétrables, il est possible d'obtenir de meilleures performances en ajustant ces paramètres pour tenir compte des spécificités des extraits de parole analysés, notamment en restreignant la plage de valeurs de f0 considérée comme plausibles en fonction du registre propre à chaque locuteur ou enregistrement. Ce constat est à l'origine de la méthode en deux étapes de De Looze (2010), qui propose d'effectuer avec l'algorithme de Praat (Boersma, 1993) une première passe de détection avec une plage de valeurs large (60Hz-600Hz), puis de procéder à une seconde passe de détection avec une plage de valeurs plus restreinte définie à partir de quantiles de la distribution des mesures obtenues lors de la première passe (minimum et maximum fixés respectivement à $0.83*Q_{15\%}$ et $1.92*Q_{65\%}$ des valeurs initiales). Si l'utilisation de valeurs limites ainsi définies permet d'améliorer sensiblement les performances comparativement à l'utilisation de valeurs par défaut, elle repose sur le postulat que la forme de la distribution obtenue lors de la première passe est comparable entre locuteurs et conditions de production.

Pour l'analyse de productions de parole et de chant dirigés vers l'enfant, Falk & Audibert (2021) ont adopté une méthodologie similaire avec une première passe de détection réalisée avec une large

plage de valeurs, mais dans laquelle les valeurs limites ont ensuite été définies pour chaque locutrice*condition à partir de l'inspection visuelle de la distribution. C'est cette dernière approche que nous illustrons à l'aide de l'outil dédié à l'affichage d'histogrammes interactifs.

2.2 Formants

Bien que la mesure automatique des formants ait fait l'objet de nombreux travaux et quand bien même l'analyse se limite à la partie supposée stable des voyelles, une telle tâche reste sujette à de nombreux biais, liés entre autres aux interactions possibles entre fréquence fondamentale et fréquence formantique (cf. Kent & Vorperian (2018) pour un récapitulatif des principaux biais). En raison de ces limites, de nombreuses études ont eu recours à une validation et correction manuelle à partir de l'inspection des spectrogrammes des relevés de fréquences formantiques effectués automatiquement (voir par exemple Van der Harst et al., 2014). Par ailleurs certains auteurs ont opté en remplacement des mesures formantiques pour une paramétrisation du signal de parole à travers des coefficients cepstraux afin de permettre l'automatisation des mesures (Ferragne & Pellegrino, 2010), ou encore pour rendre compte de variations sur les voyelles nasales pour lesquelles la détection automatique de formants est notoirement sujette à erreurs (Hermes et al., 2023), avec toutefois l'inconvénient d'une perte d'interprétabilité du lien avec l'articulation.

Une autre approche applicable à des corpus de grande taille consiste en la méthode proposée par Gendrot & Adda-Decker (2005). Il s'agit de définir un crible permettant d'éliminer les valeurs supposées aberrantes en prenant en considération la catégorie phonologique à laquelle appartient le segment analysé (généralement une voyelle) pour définir des intervalles entre lesquels la détection automatique est considérée comme correcte. La difficulté est de définir un crible suffisamment large pour ne pas assimiler à des erreurs de détection des mesures formantiques qui reflètent simplement la variabilité inhérente à la parole, notamment la parole conversationnelle susceptible de subir d'importants phénomènes de réduction segmentale. Au-delà d'erreurs grossières comme la non-détection du second formant d'où une valeur détectée excessivement élevée, le statut d'une part non-négligeable de mesures effectuées automatiquement est susceptible de rester incertain même après cette étape de filtrage. Ce constat a d'ailleurs récemment conduit Lancien et al. (2023) à opter pour une méthode statistique de filtrage fondée sur l'utilisation de distances de Mahalanobis.

Nous illustrons ici le cas de valeurs formantiques détectées automatiquement sur la voyelle /u/ pour laquelle la détection automatique de formants est problématique en raison de ses propriétés spectrales.

3 Méthodes

3.1 Données

Les données utilisées pour la mise en pratique sur des cas pratiques sont issues du corpus PTSVox (Chanclu et al., 2020), dans lequel 369 locuteurs francophones natifs ont été enregistrés dans des tâches de production de parole lue et de parole spontanée. Un sous-ensemble de 24 locuteurs (12 hommes, 12 femmes) a été enregistré dans les deux conditions lors de multiples sessions. Après une première étape de transcription manuelle et d'alignement forcé, la segmentation en mots et en phones a été corrigée manuellement pour ces 24 locuteurs. Pour les besoins de cet article, nous nous concentrons sur les voyelles produites par une locutrice âgée de 22 ans lors de l'enregistrement.

3.2 Extraction de mesures acoustiques

Les valeurs de fréquence fondamentale ainsi que les fréquences des 3 premiers formants ont été extraites au milieu de chaque voyelle, à l'aide d'un script Praat (Boersma & Weenink, 2024) développé par l'auteur. Pour l'extraction de la fréquence fondamentale, les paramètres par défaut ont été utilisés, avec un pas temporel défini automatiquement et une plage de valeurs fixée à 60-600Hz suivant les recommandations de De Looze (2010). L'extraction des fréquences formantiques a été effectuée avec l'algorithme de Burg implémenté dans Praat, avec également les paramètres par défaut (10 paramètres LPC, trames de 25ms avec recouvrement de 10ms) et une fréquence maximale pour la détection de 5 formants fixée à 5kHz pour les hommes et 5,5kHz pour les femmes.

4 iHist : histogrammes paramétrables et interactifs

4.1 Principales fonctionnalités de l'application

L'application iHist permet à l'utilisateur, après avoir importé un fichier de données au format Excel, TSV ou CSV, de sélectionner une variable quantitative à représenter sous forme d'histogramme, avec la possibilité optionnelle de filtrer les données prises en compte en fonction des valeurs d'autres variables quantitatives ou catégorielles. Parmi les principales fonctionnalités de l'application, l'utilisateur peut modifier les paramètres de l'histogramme, notamment le nombre de classes et les valeurs minimum et maximum affichées. L'application permet également d'afficher en surimpression la courbe correspondant à la densité de la distribution et/ou celle de la distribution normale de même moyenne et même écart-type. Afin de permettre à l'utilisateur de faire le lien visuellement entre la distribution dans son ensemble et les projections les plus couramment utilisées dans les représentations graphiques (graphe en barres avec erreur-type, ou boîte à moustaches), des droites verticales représentant la moyenne et l'intervalle de confiance et/ou la médiane et des quantiles sélectionnés peuvent également être ajoutés. L'application permet en outre une exploration interactive des données de l'histogramme, en accédant au détail des observations regroupées dans une classe via un clic sur la barre correspondante. De cette façon, l'utilisateur peut afficher ces informations, et les exporter dans un fichier structuré pour simplifier leur inspection ultérieure.

La version française de cette application est disponible en ligne à l'URL suivant : https://shiny.laboratoirephonetiquephonologie.fr/iHist_fr/ . Le code source de l'application ainsi que les fichiers de localisation en français et en anglais sont distribués sous licence GPL via GitHub[1].

4.2 Application aux valeurs de f0

Sur les mesures de f0 obtenues à partir des 4 972 voyelles produites par la locutrice sélectionnée, les valeurs limites obtenues pour la seconde passe de détection avec la méthode de De Looze (2010) sont de $0,83*Q_{15\%}$=131Hz et de $1,92*Q_{65\%}$=303Hz. Pour cette illustration (Figure 1) nous nous concentrons sur la limite basse, sous laquelle on trouve un nombre de valeurs plus important qu'attendu comme le montre l'histogramme. Si l'inspection de la première classe (70,4Hz-96,2Hz) confirme que les valeurs de f0 s'y trouvant correspondent bien à des erreurs de détection, les valeurs

[1] https://github.com/nicolasaudibert/iHist.git

proches de cette limite basse nécessitent une inspection plus approfondie afin d'affiner cette valeur limite, et le cas échéant définir des sous-groupes. Cette inspection est ici réalisée via l'exportation au format TSV des valeurs sélectionnées après affinage du nombre de classes, utilisée ensuite comme entrée d'un script Praat dédié à l'affichage et l'évaluation d'extraits sélectionnés.

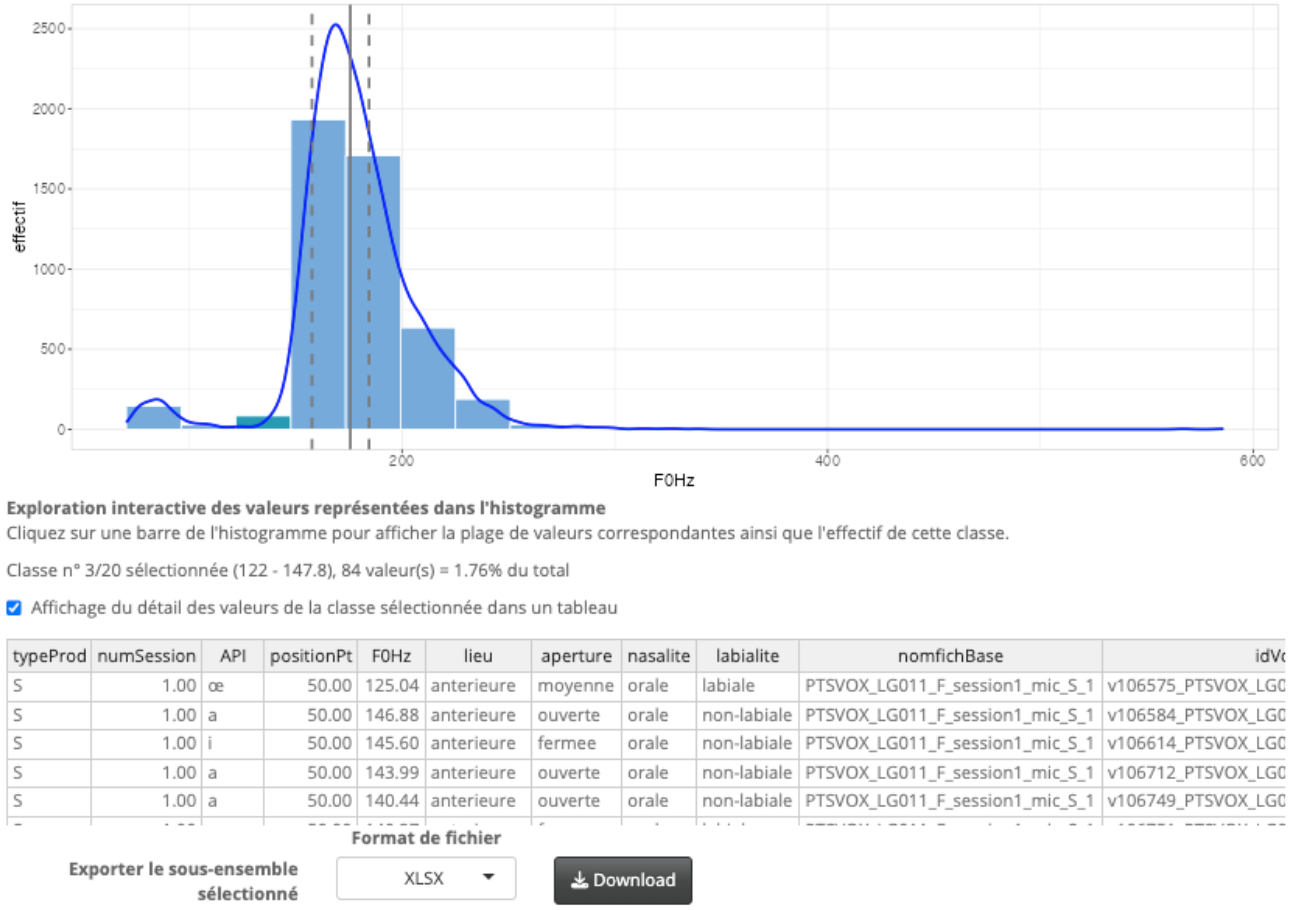


FIGURE 1 : Illustration de l'application *iHist* appliquée aux valeurs de f0 de la locutrice LG011, après sélection d'une classe (en vert) contenant 1,76% de l'ensemble des valeurs dont le détail est affiché dans le tableau en bas de page. Outre l'adaptation du nombre de classes affichées, l'affichage de la densité de distribution (courbe bleue) est activé, ainsi que celui de la médiane et de quantiles sélectionnés (droites grises, ici en pointillé quantiles à 15% et 65%).

5 iScatter : nuages de points interactifs

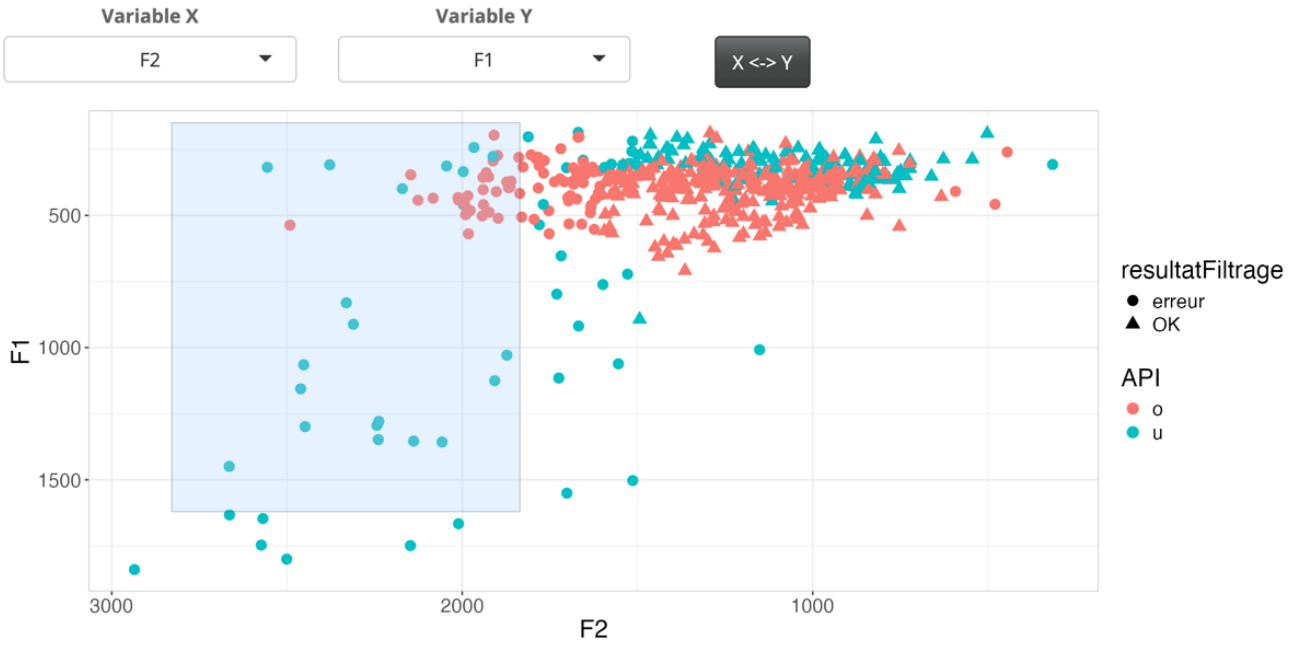
5.1 Principales fonctionnalités de l'application

Outre les fonctionnalités d'importation et de filtrage des données équivalentes à celles de iHist, l'application iScatter dédiée à l'affichage de nuages de points interactifs permet à l'utilisateur de sélectionner une ou deux variables indépendantes qui définissent la couleur et/ou la forme des points, et de façon optionnelle d'afficher les droites de régression pour chaque sous-groupe défini par les variables indépendantes. Les corrélations de Pearson et Spearman sont en outre affichées pour chaque sous-groupe. L'utilisateur a la possibilité de sélectionner le point le plus proche par un clic ou d'effectuer une sélection rectangulaire pour afficher les détails des points sélectionnés dans un

tableau exportable. La sélection d'une ligne dans ce tableau permet de visualiser la position du point correspondant dans le nuage de points.

La version française de l'application iScatter est disponible en ligne à l'URL suivant : https://shiny.laboratoirephonetiquephonologie.fr/iScatter_fr/. Le code source de l'application et les fichiers de localisation sont également distribués sous licence GPL via GitHub[2].

5.2 Application aux valeurs de formants : cas de la voyelle /u/



Affichage du détail des valeurs sélectionnées dans un tableau

loc	sexeLoc	typeProd	duree	API	ctxG_API	ctxD_API	F1	F2	F3	F4	resultatFiltrage
LG011	F	S	0.02	o	j	n	481.46	1977.24	3302.63	4323.47	erreur
LG011	F	S	0.02	u	t	ʒ	911.89	2310.52	3511.31	4428.10	erreur
LG011	F	S	0.04	u	t	ʃ	1348.10	2239.53	3840.36	4665.96	erreur
LG011	F	S	0.04	o	s	m	389.39	1864.96	2868.84	4674.40	erreur
LG011	F	S	0.06	o	n	k	392.51	1863.79	3104.94	4076.96	erreur

FIGURE 2 : Illustration de l'application *iScatter* appliquée aux valeurs de formants de la locutrice LG011, après application d'un filtre pour n'afficher que les occurrences de /u/ et /o/ et sélection rectangulaire d'une zone incluant une majorité de valeurs étiquetées comme erronées suite à l'application d'un crible. La première variable indépendante (couleurs distinctes) est utilisée pour distinguer les catégories vocaliques, et la seconde (formes des points) pour distinguer les exemplaires étiquetés par le crible comme ayant des valeurs formantiques correctes ou non

La figure 2 illustre l'utilisation de l'application iScatter pour explorer les valeurs de F1 et F2 considérées comme erronées, suite à l'application d'un crible avec les mêmes valeurs limites que celles utilisées par Audibert et al. (2015) sur les mesures formantiques des 4 255 voyelles orales produites par la locutrice LG011. La sélection effectuée pour afficher les détails des caractéristiques de certaines voyelles se concentre ici sur les occurrences de la voyelle /u/ pour lesquelles les valeurs de F2 détectées automatiquement dépassent la limite supérieure de 1500Hz définie par le crible

2 https://github.com/nicolasaudibert/iScatter.git

utilisé pour les /u/ produits par des femmes (par ailleurs pour de nombreux exemplaires de /u/, la valeur de F1 détectée dépasse également la limite supérieure de 1000Hz).

Notons ici que parmi les options de paramétrage de l'affichage du nuage de point proposées par l'application, l'inversion de l'orientation des axes *x* et *y* a été utilisée afin d'obtenir une représentation conforme à celle couramment utilisée pour représenter les voyelles dans le plan F1/F2. Par ailleurs afin d'éviter d'afficher un nombre de points superposés trop important, un filtre a été appliqué afin de n'afficher que les occurrences de /u/, ainsi qu'à titre de comparaison les occurrences de la voyelle /o/ avec laquelle le recouvrement est particulièrement important dans les données de cette locutrice. De même qu'avec l'application iHist, en complément de l'inspection immédiate des caractéristiques documentées dans le jeu de données importé, le sous-ensemble sélectionné peut-être exporté pour faire l'objet d'une inspection guidée des signaux de parole (en l'occurrence également à l'aide d'un script Praat).

6 Discussion et conclusion

Les cas d'utilisation exposés dans cet article se concentrent sur l'utilisation des outils proposés pour identifier des erreurs de mesure. Cependant ces cas d'utilisation constituent une simple illustration, ces outils pouvant être utiles pour de nombreuses autres applications potentielles liées à l'étude de la voix et de la parole. En particulier, l'application *iScatter* dédiée aux nuages de points interactifs utilisable pour identifier les observations qui ne suivent pas la tendance générale dans le cadre de l'exploration des liens entre variables quantitatives, plus spécifiquement les relations entre mesures acoustiques et/ou perceptives supposées capturer différentes facettes d'un même phénomène. À ce titre, cette application est déjà utilisée régulièrement par des collègues pour l'inspection de données de voix et de parole, mais n'avait jamais fait l'objet d'une diffusion plus large dans la communauté.

Sans avoir la prétention de fédérer une communauté autour du développement de ces outils dont de nombreux aspects restent perfectibles, la distribution de leur code source sous licence GPL permet à tout un chacun de les modifier pour les adapter à des besoins spécifiques, et surtout de les utiliser localement avec des fichiers plus volumineux que ne le permet le serveur utilisé pour rendre ces applications directement accessibles en ligne. En effet, leur utilisation sur un ordinateur personnel nécessite uniquement l'installation des logiciels R et RStudio (complétés par un ensemble de paquets) auxquels la majorité des collègues et étudiants amenés à procéder à l'analyse quantitative de données de voix et de parole sont déjà familiarisés.

La vocation des outils proposés est avant tout d'offrir aux collègues et étudiants un complément à l'utilisation de tableurs pour l'exploration des données quantitatives, afin de faciliter la mise en œuvre de l'injonction fréquente dans les formations en statistiques quel que soit le niveau visé : « N'oubliez pas de regarder vos données ! ». En revanche il ne s'agit en aucun cas de se substituer aux formations à l'utilisation de R et autres outils avancés d'analyse de données, qui restent indispensables et doivent continuer à être encouragées.

Remerciements

Ce travail a été soutenu par le projet ANR PASDCODE (ANR-21-CE28-0015) et par le Laboratoire d'Excellence Empirical Foundations of Linguistics (LabEx EFL, ANR-10-LABX-0083). Il contribue à l'IdEx Université de Paris (ANR-18-IDEX-0001).

Références

AUDIBERT, N., FOUGERON, C., GENDROT, C., & ADDA-DECKER, M. (2015). Duration-vs. style-dependent vowel variation: A multiparametric investigation. In *Proceedings of the 18th International Congress of Phonetic Sciences*, p. 5. HAL: hal-01251372.

CHANG W., CHENG J., ALLAIRE J., SIEVERT C., SCHLOERKE B., XIE Y., ALLEN J., MCPHERSON J., DIPERT A. & BORGES B. (2024). *shiny: Web Application Framework for R*. R package version 1.8.0.9000, https://github.com/rstudio/shiny, https://shiny.posit.co/.

BOERSMA P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences* (Vol. 17, No. 1193, pp. 97-110).

BOERSMA P. & WEENINK D. (2024). Praat: doing phonetics by computer [Programme informatique]. Version 6.4.05, téléchargé le 27 janvier 2024 depuis http://www.praat.org/

CHANCLU A., GEORGETON L., FREDOUILLE C. & BONASTRE J. F. (2020). PTSVOX: une base de données pour la comparaison de voix dans le cadre judiciaire. In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1 : Journées d'Études sur la Parole* (pp. 73-81). HAL: hal-02798519.

DE LOOZE C. (2010). Analyse et Interprétation de l'Empan Temporel des Variations Prosodiques en Français et en Anglais. Thèse de doctorat. Université de Provence - Aix-Marseille I. HAL: tel-00470641.

FALK S. & AUDIBERT N. (2021). Acoustic signatures of communicative dimensions in codified mother-infant interactions. *The Journal of the Acoustical Society of America*, *150*(6), 4429-4437. DOI: 10.1121/10.0008977. HAL: hal-03592269.

FERRAGNE E. & PELLEGRINO F. (2010). Vowel systems and accent similarity in the British Isles: Exploiting multidimensional acoustic distances in phonetics. *Journal of Phonetics*, *38*(4), 526-539. DOI: 10.1016/j.wocn.2010.07.002. HAL: hal-01240095.

GENDROT C. & ADDA-DECKER, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. In *Proceedings of Interspeech 2005*, p. 2453-2456. DOI: 10.21437/Interspeech.2005-753. HAL: halshs-00188096.

HERMES A., AUDIBERT N. & BOURBON A. (2023). Age-related vowel variation in French. In *Proceedings of the 20th International Congress of Phonetic Sciences*, p. 2045-2049. Guarant International. HAL: hal-04193397.

JOUVET D. & LAPRIE Y. (2017). Performance analysis of several pitch detection algorithms on simulated and real noisy speech data. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, p. 1614-1618. IEEE. DOI: 10.23919/EUSIPCO.2017.8081482. HAL: hal-01585554.

KENT R. D. & VORPERIAN H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of communication disorders*, *74*, 74-97. DOI: 10.1016/j.jcomdis.2018.05.004.

LANCIEN M., ADDA-DECKER M. & STUART-SMITH J. (2023). Knowledge-driven vs. data-driven methods for filtering acoustic measures in phonetics corpora. In: Skarnitzl R. & Volín J. (Eds.), In *Proceedings of the 20th International Congress of Phonetic Sciences*, p. 3166-3170. Guarant International.

LIBERMAN M. Y. (2019). Corpus phonetics. *Annual Review of Linguistics*, *5*, 91-107. DOI: 10.1146/annurev-linguistics-011516-033830.

NICKLIN C. & PLONSKI L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics*, *40*, 26-55. DOI: 10.1017/S0267190520000057.

OSBORNE J. W. & OVERBAY A. (2019). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, *9*(1), 6. DOI: 10.7275/qf69-7k43.

R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

VAN DER HARST S., VAN DE VELDE H. & VAN HOUT R. (2014). Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change*, *26*(2), 247-272. DOI: 10.1017/S0954394514000040.

VAYSSE R., ASTÉSANO C. & FARINAS J. (2022). Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech. *The Journal of the Acoustical Society of America*, *152*(5), 3091-3101. DOI: 10.1121/10.0015143. HAL: hal-03879676.

WICKHAM H., AVERICK M., BRYAN J., CHANG W., MCGOWAN L.D., FRANÇOIS R., GROLEMUND G., HAYES A., HENRY L., HESTER J., KUHN M., PEDERSEN TL., MILLER E., BACHE S.M., MÜLLER K., OOMS J., ROBINSON D., SEIDEL D.P., SPINU V., TAKAHASHI K., VAUGHAN D., WILKE C., WOO K. & YUTANI H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, **4**(43), 1686. DOI: 10.21105/joss.01686.