

“Le voyage d’Alice” - Extraction automatique de mesures acoustiques de la parole et de la voix sur une lecture à voix haute du texte standardisé

Contributeurs & contact

- **Timothy Pommée** : timothy_pommee@hotmail.com
<https://orcid.org/0000-0001-7846-7282>
- **Liziane Bouvier**
<https://orcid.org/0000-0002-7239-7639>
- **Julien Pinquier**
<https://orcid.org/0000-0003-1556-1284>
- **Julie Mauclair**
<https://orcid.org/0000-0002-2740-5118>
- **Véronique Delvaux**
<https://orcid.org/0000-0002-4200-2982>
- **Cécile Fougeron**
<https://orcid.org/0000-0002-2166-4602>
- **Corine Astésano**
<https://orcid.org/0000-0002-0882-4974>
- **Vincent Martel-Sauvageau**
<https://orcid.org/0000-0003-0152-0908>
- **Dominique Morsomme**
<https://orcid.org/0000-0002-7697-0498>
- **Muriel Lalain**
<https://orcid.org/0000-0002-7672-8589>
- **Virginie Woisard**
<https://orcid.org/0000-0003-3895-2827>
- **Pierre Pinçon**

Publication

Pommée, T., Pinquier, J., Mauclair, J., Bouvier, L., Delvaux, V., Fougeron, C., Astésano, C., Martel-Sauvageau, V., Morsomme, D., Pinçon, P., Lalain, M., & Woisard, V. (2023). **Le voyage d’Alice : un texte standardisé pour l’évaluation de la parole et de la voix en Français**. *Glossa*, XXX.

Introduction

Le texte “Le voyage d’Alice” a été créé spécifiquement pour l’évaluation de la parole et de la voix en français, sur base d’un ensemble exhaustif de critères, prenant en compte les données de la littérature, les besoins spécifiques identifiés en recherche scientifique et en pratique clinique francophone, ainsi que les données d’une étude de consensus internationale. Sa construction est décrite en détails et en français dans la publication dont la référence est renseignée ci-dessus.

Ce texte est destiné à fournir un support standardisé pour l’évaluation :

- de l'articulation des sons de la parole (dysarthrie, apraxie) ;
- des variations prosodiques et du comportement phonatoire (dysphonie, harmonisation vocale) ;
- de la fluence/des disfluences (bégaiement/bredouillement) ;

chez les locuteurs âgés d'au moins 12 ans.

Il s'est montré utilisable et facilement lisible en Belgique, en France comme au Canada.

Le texte standardisé

Le voyage d'Alice

Lundi matin, Alice et son Papa vont à Malibou.

Là-bas, ils rejoignent Papy après un voyage sans soucis.

Il fait chaud, mais la brise légère et l'air iodé de la mer les ravivent.

Vers midi, Alice s'exclame : « J'ai vraiment très très faim ! ».

Papy les guide alors vite vers un café luxueux au bord de l'eau : Le Bigorneau Salé.

Mardi, ils vont à la plage.

Il n'y a pas un nuage dans le ciel. Papa s'interroge : « Avons-nous pris la crème solaire ? »

« Bien sûr ! », répond Alice.

Mercredi, Papa et Papy se baladent en bavardant.

Pendant ce temps, Alice se détend en lisant un roman et mange un bonbon à l'ananas.

Jeudi, elle va faire un jogging.

Papa lui crie : « Nous partons faire quelques achats ! »

Au magasin, Papy achète des noix de macadamia.

Vendredi, ils visitent un musée d'art abstrait.

Papa s'extasie devant un splendide tableau et demande : « Qui a donc créé cette œuvre ? ».

Samedi matin, Alice s'entraîne pour la soirée karaoké en répétant rapidement : « pataka pataka pataka ».

Samedi soir, ils fêtent leur départ en dansant la java sous le lilas.

Comme à l'arrivée, il fait chaud, mais la brise légère et l'air iodé de la mer les ravivent.

Dimanche, Alice, Papa et Papy quittent Malibou.

Ils rentrent affamés.

À table, il y a de la pizza garnie et des lasagnes aux champignons.

Rassasiés, ils s'exclament : « Quel séjour extraordinaire ! »

Outil automatique d'analyse acoustique

Le texte étant destiné à l'utilisation par des chercheurs et des cliniciens d'horizons variés, nous avons rédigé un protocole d'extraction de mesures acoustiques libre d'accès et facile d'utilisation par des professionnels non experts en informatique. Celui-ci permet le prétraitement des fichiers audio, un alignement forcé (p. ex., en utilisant le service gratuit en ligne « WebMAUS Basic »¹), puis l'extraction de mesures de qualité vocale et de mesures liées à l'intelligibilité de la parole pour chacune des trois classes de phonèmes (voyelles, consonnes, semi-consonnes) :

- 1) Mesures de **qualité vocale**² sur les deux occurrences de la phrase voisée (« mais la brise légère et l'air iodé de la mer les ravivent ») :

¹ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

² Adapté du script de l'Acoustic Voice Quality Index version 03.01 [Barsties 16]

- **la proéminence du pic cepstral lissé** (« CPPs ») : le cepstre est « une représentation spectrale du spectre » [Heman-Ackah 03]. Le pic proéminent sur le cepstre est appelé pic cepstral, et il est réciproque à la distance séparant les harmoniques du spectre (il correspond ainsi à la période fondamentale du son) [Hillenbrand 96]. Plus le pic cepstral est prononcé, plus le spectre sonore est périodique. Ainsi, plus une voix est dysphonique, moins la composition spectrale harmonique sera « propre », et moins le pic sera discernable. Le CPPs correspond à la distance entre l'amplitude du pic cepstral et l'amplitude du point ayant la même quéfrencce sur la droite de régression à travers le cepstre lissé³ (donc l'amplitude prédite à ladite quéfrencce par la droite de régression, cf. Figure 1) [Hillenbrand 94]. L'utilisation de données du domaine cepstral permet des mesures plus fiables dans les dysphonies sévères [Heman-Ackah 03]. En effet, ces dernières sont caractérisées par un signal peu périodique et se prêtent plus difficilement aux analyses du domaine fréquentiel tels le shimmer, le jitter et le rapport signal/bruit. De plus, les mesures cepstrales permettent l'évaluation du degré de la dysphonie sur un échantillon de parole continue [Delgado-Hernández 18; Lowell 11; Sauder 17] ;

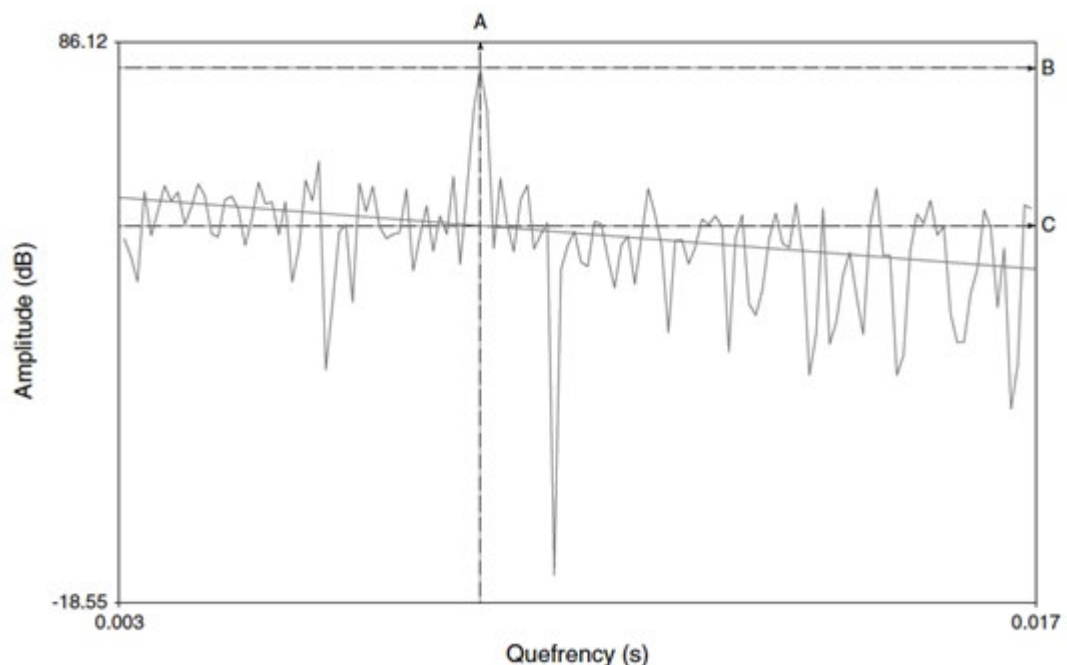


Figure 1. Représentation du pic cepstral (Maryn, De Bodt, Barsties, & Roy, 2014). La proéminence du pic cepstral (A) est calculée en mesurant la différence entre le sommet de celui-ci (B) et le point de même quéfrencce sur la droite de régression (C).

³ Le CPPs est mesuré après un « lissage », c'est-à-dire un moyennage d'un certain nombre de fenêtres cepstrales avant et après la fenêtre d'intérêt [Heman-Ackah 03; Hillenbrand 96].

- **la pente du spectre moyen à long terme** (« LTAS slope ») : Le spectre moyen à long terme (long-term average spectrum — LTAS en anglais) est un spectre obtenu en moyennant un grand nombre de spectres à intervalles réguliers (p.ex. toutes les 200 ms, donc cinq spectres/seconde) (cf. Figure 2) [Master 06]. Les mesures du spectre moyen à long terme (tout comme les mesures cepstrales) ont gagné de l'intérêt dans l'évaluation vocale acoustique, car elles permettent tout comme les mesures cepstrales l'évaluation de la sévérité de la dysphonie sur des échantillons de parole continue [Lowell 11]. L'intensité des différents harmoniques d'un son décroît avec l'augmentation des fréquences ; ainsi, la « pente spectrale » (exprimée en dB), reliant les points d'amplitude maximale entre les harmoniques, chute. On parle généralement d'une chute de 12 dB par octave en voix « normale ». La pente du spectre moyen à long terme se mesure en divisant l'énergie des fréquences basses (0-1 kHz) par celle des fréquences hautes (1-10 kHz) à travers le spectre moyen à long terme [Volín 11]. Une énergie harmonique diminuée dans les hautes fréquences (donc une pente accentuée) peut être liée à une voix soufflée ou hypofonctionnelle [Hammarberg 80 ; Dromey 03];
- **l'inclinaison de la courbe de tendance à travers le spectre moyen à long terme** (« LTAS tilt ») : il s'agit de la différence d'énergie de la droite de régression (prédiction de l'énergie en fonction de la fréquence, cf. Figure 2) entre 0-1 kHz et 1-10 kHz à travers le spectre moyen à long terme [Barsties 15; Smith 14] ;

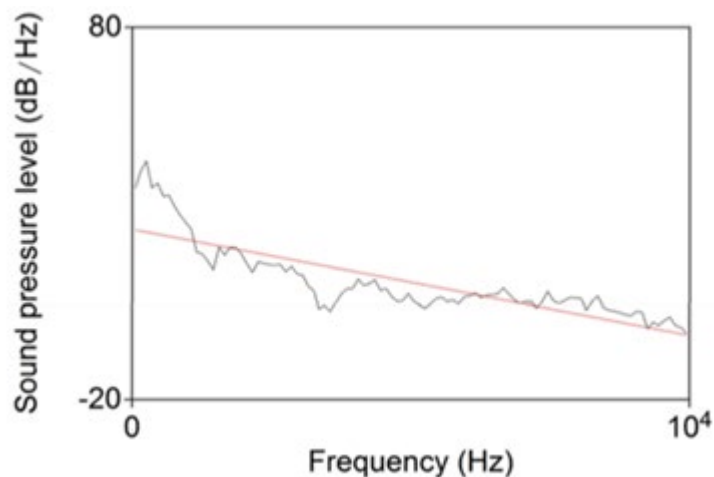


Figure 2 - Représentation du « spectral tilt » [Delgado-Hernandez 17] ; en noir = spectre moyen à long terme ; en rouge = droite de régression à travers le spectre moyen à long terme

- 2) **Mesures vocaliques**⁴ : extraction des deux premiers formants des voyelles cardinales, traçage du **triangle vocalique** et calcul de l'**aire du triangle** relative à un

⁴ Adapté du script VowelTriangle (<https://github.com/robvanson/VowelTriangle>) [Van Son 18]

triangle de référence. Un exemple de triangle vocalique obtenu à l'aide du script employé est illustré en Figure 3. Pour plus de détails quant à sa construction, voir [Van Son 18]. La mesure VSA est sensible aux différences d'intelligibilité dans la parole saine [Bond 94] et est couramment utilisée pour rendre compte de la centralisation des voyelles [Liu 05; Sapir 09; Weismer 01].

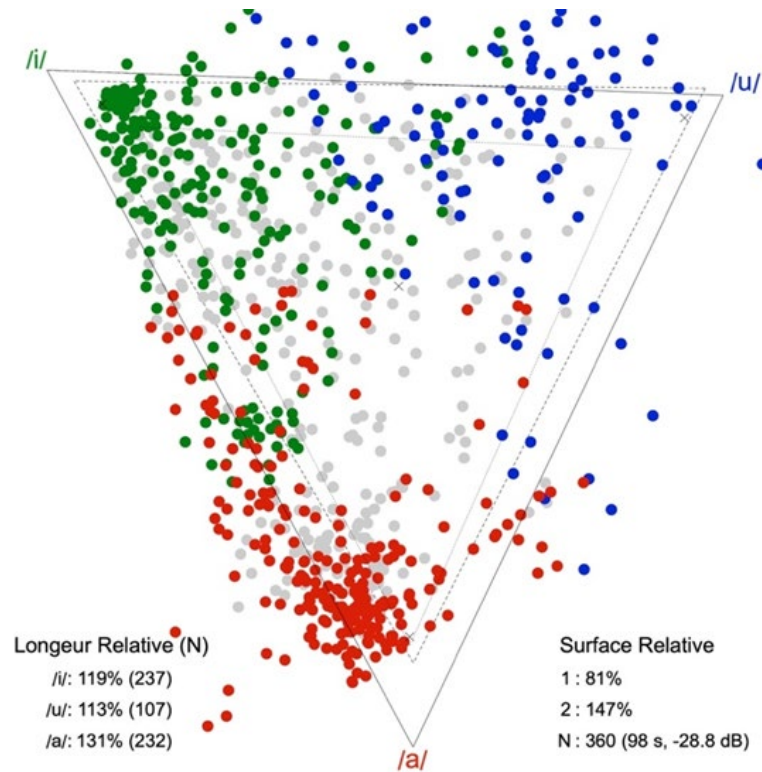


Figure 3 – Exemple de triangle vocalique obtenu de la lecture du texte entier par une locutrice belge. F1 en ordonnées (haut-bas), F2 en abscisses (droite-gauche), en demi-tons ; ligne discontinue : triangle de référence ; pointillés : triangle employé par le locuteur (1 E.T.) ; ligne continue : triangle employé par le locuteur (2 E.T.) ; « x » : voyelles cardinales et centroïde (moyenne géométrique des fréquences des voyelles cardinales) ; 1 = surface relative du triangle 1 E.T. ; 2 = surface relative du triangle 2 E.T. ; N = nombre total de segments vocaliques

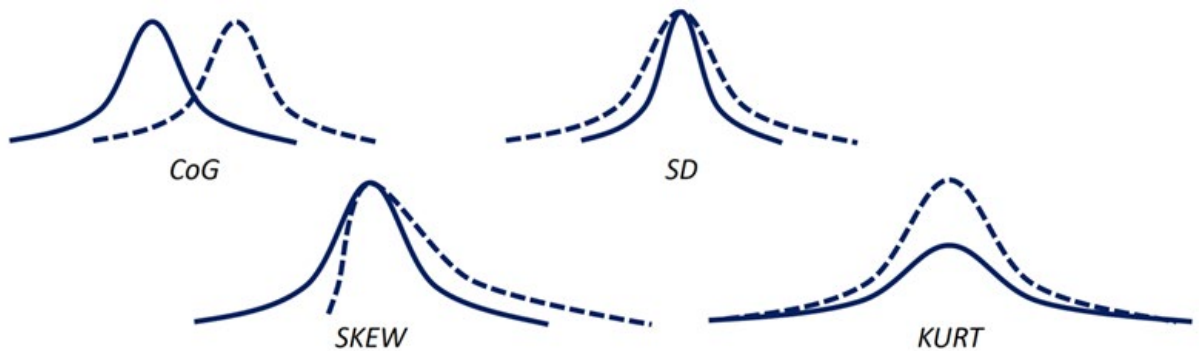
3) **Mesures consonantiques** : calcul des **moments spectraux** sur les consonnes en contexte /aCa/. Lors de l'analyse en moments spectraux, le spectre consonantique est considéré comme une distribution statistique et est décrit par quatre mesures [Jongman 00] (voir Figure 4) : le centre de gravité (« center of gravity » ou CoG), l'écart-type (« standard deviation » ou SD), l'asymétrie (« skewness », SKEW) et l'aplatissement (« kurtosis », KURT). Le **premier moment (CoG, en Hertz)** est la « fréquence qui divise le spectre en deux moitiés de telle sorte que la quantité d'énergie dans les régions de haute fréquence est égale à celle dans la région de basse fréquence » [Yoon 15]. Par conséquent, un CoG élevé indique que le noyau spectral est situé dans les hautes fréquences, par exemple dans les sibilantes

[Hernandez 19]. Le **second moment (SD)**, en Hertz, mesure la dispersion du noyau spectral autour de ce centre de gravité. Un petit SD indique une concentration étroite autour du CoG. Le **troisième moment (SKEW)** est généralement considéré comme étant lié au CoG [Cattelain 19; Newman 03; Tjaden 10]. Alors qu'une asymétrie de 0 correspond à une distribution symétrique, une valeur positive indique que la distribution est asymétrique vers la droite, c.-à-d. que la queue droite de la distribution (valeurs au-dessus du mode et de la moyenne) s'étend plus loin que la queue gauche [Jongman 00]. SKEW peut également être interprété comme faisant référence à l'inclinaison spectrale [Jongman 00], une valeur positive indiquant une inclinaison négative (une forte concentration spectrale dans les basses fréquences), et inversement. Le **quatrième moment (KURT)** est quant à lui lié au SD [Cattelain 19; Newman 03; Tjaden 10] et constitue une mesure de l'acuité du pic. Plus le KURT est élevé, plus le pic est défini. Notons que SKEW et KURT ont tous deux des unités sans dimension [Gottsmann 11].

Il est communément admis que la forme spectrale est caractéristique de la longueur et de la forme de la cavité antérieure à la constriction articulatoire [Behrens 88; Kay 12; Stevens 98; Yoon 15]. Ainsi, pour les consonnes plus antérieures, le noyau spectral est concentré dans les fréquences plus élevées. De même, plus la cavité antérieure est longue, meilleure sera la définition de la forme spectrale [Stevens 98]. De ce fait, les fricatives labiodentales et plosives labiales présentent en théorie des spectres plutôt plats, alors que les consonnes (palato-) alvéolaires présentent une forme spectrale bien définie et pointue [Jongman 00; Repp 89]. Des liens directs entre les moments spectraux et les mesures articulatoires physiologiques ont été suggérés, par exemple entre le CoG et le SKEW dans les plosives et le mouvement de la langue [Cattelain 19], ainsi qu'entre le CoG dans les fricatives et les mouvements de la mâchoire [Iskarous 11]. De ce fait, ces mesures suscitent entre autres un intérêt en application clinique [Acher 14; Hernandez 19; Kay 12; Kent 99; Tjaden 17].

De nombreuses études montrent que les moments spectraux peuvent être utilisés pour « objectiver » les différences entre la parole naturelle et « atypique ». En outre, leur mesure est non invasive et peut être effectuée avec peu d'équipement [Meyer 11]. Ainsi, les moments spectraux ont pu être utilisés pour évaluer la dysarthrie (p. ex. [Dromey 03; Hernandez 19; Kay 12; Kent 99; Lam 16; Tjaden 97, 17]). Ils ont également été appliqués pour des patients glossectomisés (p. ex. [Acher 14]), des personnes malentendantes (p. ex. [Abraham 20; Blacklock 04; Scarbel 12]), mais aussi le développement de la parole chez les enfants (p. ex. [Körkkö 15; Munson 04; Nissen

03, 05]). Ils seraient donc *a priori* pertinents pour l'évaluation de l'efficacité thérapeutique ou de la progression pathologique [Tjaden 10].



- 4) **Mesure semi-consonantique** : mesure de la **pen**te des trois premiers formants dans les semi-consonnes. Le script que nous avons écrit permet d'extraire les formants selon la méthode de Burg sur chaque fenêtre de 5 ms du début à la fin du segment semi-consonantique. Le calcul des pentes se fait ensuite en soustrayant la moyenne des 5 dernières fenêtres de la moyenne des 5 premières fenêtres, et en divisant le tout par la durée du segment (en ms). Dans le vieillissement sain, une diminution de l'intelligibilité peut être en partie attribuée à un ralentissement des mouvements de la langue [Kuruvilla-Dugdale 20]. En parole pathologique, la pente de F2 est considérée comme un indicateur de la lenteur articulatoire [Kent 89; Y. Kim 09; Weismer 92, 95].

Consignes d'installation

Pré requis:

ATTENTION: Ce script ne fonctionne pour l'instant que sur Mac et Linux. Nous travaillons sur son adaptation pour Windows.

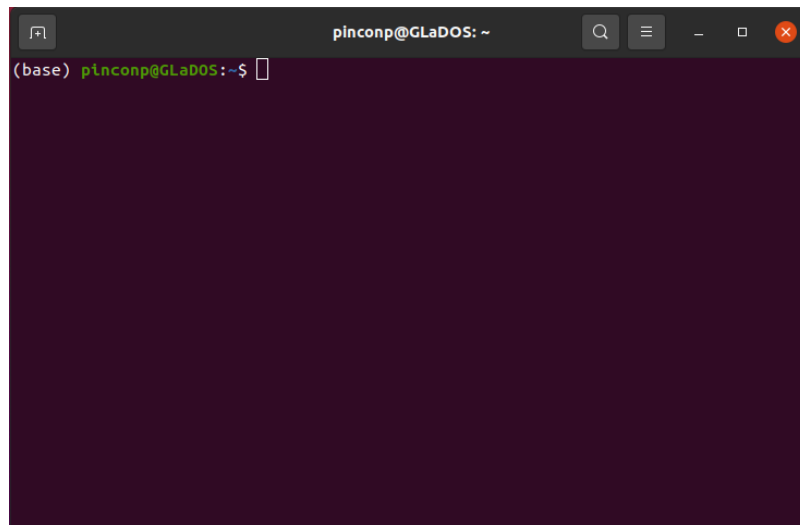
A. Praat: téléchargez la dernière version de Praat via: <https://www.fon.hum.uva.nl/praat>

B. Conda:

1. Rendez-vous sur le site officiel de Conda :
<https://conda.io/projects/conda/en/latest/user-guide/install/index.html>.
2. Cliquez sur le bouton "Télécharger" correspondant à votre système d'exploitation (Linux ou macOS).
3. Suivez les instructions fournies par le programme d'installation. Sous Linux, vous avez le choix entre l'installateur Anaconda et l'installateur Miniconda, tous deux fournis sous forme de scripts shell. Sous macOS, vous devez télécharger un fichier d'installation PKG.
4. Une fois l'installation terminée, vous pouvez ouvrir une nouvelle fenêtre de terminal pour tester si Conda a été installé correctement. Exécutez la commande suivante :
"conda --version"
Si Conda est correctement installé, le terminal affichera le numéro de version de Conda.

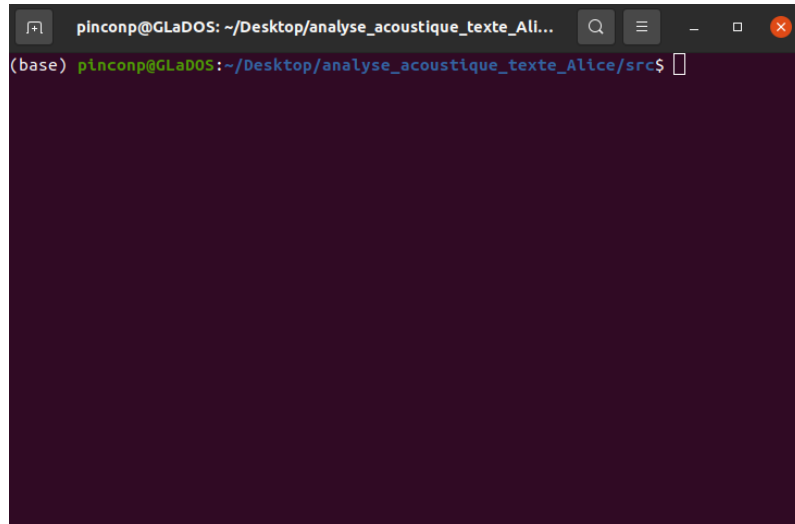
Attention, il ne faut pas oublier de fermer le terminal puis en ouvrir un nouveau pour que l'installation prenne effet.

Si conda est bien installé, vous devriez voir "(base)" au début de la ligne de commande de votre terminal.



Pour l'installation spécifique au script :

1. Ouvrez un terminal depuis le sous-dossier "analyse_acoustique_texte_Alice/src".
C'est par ce chemin d'accès que devrait finir la ligne de commande du terminal.

A terminal window with a dark purple background. The title bar at the top reads "pinconp@GLaDOS: ~/Desktop/analyse_acoustique_texte_Ali...". The terminal content shows a shell prompt "(base) pinconp@GLaDOS: ~/Desktop/analyse_acoustique_texte_Alice/src\$" followed by a cursor. The rest of the terminal is empty.

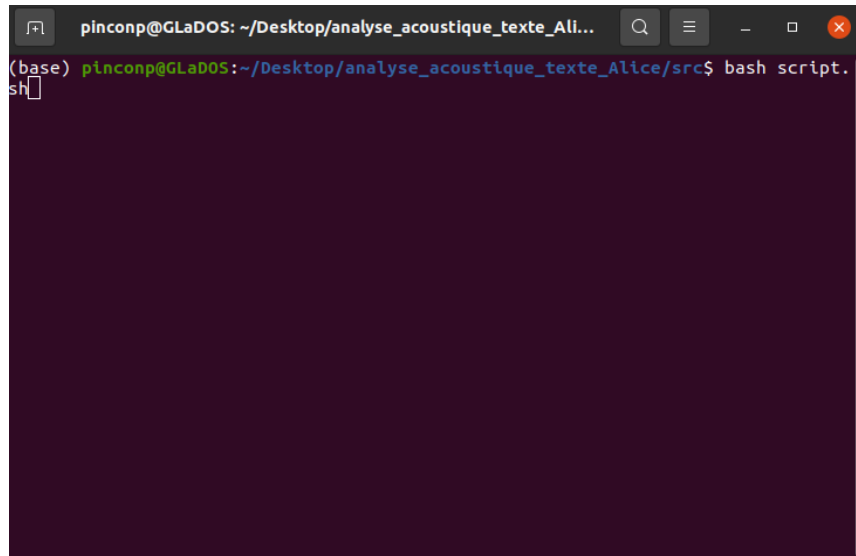
```
pinconp@GLaDOS: ~/Desktop/analyse_acoustique_texte_Ali...
(base) pinconp@GLaDOS: ~/Desktop/analyse_acoustique_texte_Alice/src$
```

2. Tapez "bash install.sh", puis rentrez votre mot de passe lorsque cela l'est demandé. L'installation prend quelques minutes.

Consignes d'utilisation

Le script doit être utilisé comme suit :

1. Lancer le script.



```
pinconp@GLaDOS: ~/Desktop/analyse_acoustique_texte_Ali...
(base) pinconp@GLaDOS:~/Desktop/analyse_acoustique_texte_Alice/src$ bash script.
sh
```

2. Le nom des fichiers à analyser doit suivre le formalisme suivant :
"nom_numéroenregistrement_genre.wav".
3. Placer les enregistrements tel que demandé dans le terminal. A savoir, placer les fichiers masculins dans
"analyse_acoustique_texte_Alice/src/1_wav_originaux/FR/Hommes" et les fichiers féminins dans "analyse_acoustique_texte_Alice/src/1_wav_originaux/FR/Femmes"
4. Si des fichiers d'alignement Paty existent (optionnel), les placer dans le dossier "analyse_acoustique_texte_Alice/src/Paty_alignment", sinon un alignement moins précis en ligne sera effectué.
5. Une fois ces étapes effectuées, retourner sur le Terminal et appuyer sur "Entrée".
6. Indiquer en tapant "1" ou "2" dans le Terminal si c'est une lecture du texte complet ou la version courte. Le traitement prend quelques minutes.
7. Les résultats se trouvent dans le dossier
"analyse_acoustique_texte_Alice/src/Analyzed_results".

Recommandations générales pour des enregistrements de qualité

Recommandations - conditions d'enregistrement

[Chial 03; Deliyki 05; R. R. Patel 18; Plichta 02; Šrámková 15; Švec 10, 18; Winholtz 97]

À des fins de recherche, tous les paramètres suivants doivent demeurer constants ; à des fins cliniques, les paramètres doivent demeurer constants pour chaque patient (p. ex. le niveau d'entrée peut parfois nécessiter une adaptation en fonction du profil du patient, comme pour les chanteurs ou pour les patients ayant un niveau d'intensité vocale très bas, mais il doit ensuite être réglé de manière identique pour chaque enregistrement suivant de ces mêmes patients).

Suggestions générales :

- Essayez de vous en tenir à un seul logiciel d'enregistrement/d'analyse : Praat (gratuit), Audacity (gratuit), Computerized Speech Lab (commercial)...
- Calibrez le microphone pour les mesures d'intensité : enregistrez un son pur stable provenant d'un haut-parleur ou une voyelle tenue avec le microphone de type serre-tête et utilisez simultanément un sonomètre à la même distance ; comparez l'intensité mesurée par le microphone (par exemple en utilisant Praat) avec la valeur du sonomètre, qui vous donne la valeur de référence. Si, par exemple, Praat indique 65 dB SPL alors que le sonomètre indique 60 dB SPL, pour toutes les mesures ultérieures effectuées avec exactement la même distance et les mêmes paramètres, soustrayez 5 dB de la mesure de l'ordinateur pour obtenir le niveau de pression acoustique « réel ».
- Effectuez les enregistrements dans une pièce calme. Si aucune pièce insonorisée n'est disponible, mesurez le bruit ambiant à l'aide d'un sonomètre ; en considérant que le rapport signal/bruit doit être supérieur à 42 dB pour des mesures acoustiques fiables (bien que ≥ 30 dB soit acceptable), et que l'intensité moyenne de la voix conversationnelle est d'environ 60 dB, le bruit ambiant doit être inférieur à 18 dB pour permettre un rapport signal/bruit optimal.
- Évitez les sources de bruit telles que les ventilateurs/climatisations, les souffleries d'ordinateurs, les bourdonnements de néons... mais aussi les smartphones et autres accessoires connectés comme les montres connectées, qui peuvent provoquer des interférences sonores.

Microphone :

- microphone de type serre-tête pour assurer une distance constante entre le microphone et la bouche et maximiser le rapport signal/bruit
- microphone cardioïde (directionnel) pour limiter l'effet du bruit ambiant et maximiser le rapport signal/bruit (à des fins de recherche, microphone omnidirectionnel pour une meilleure précision en termes de niveau de pression sonore et de fréquence)
- type de transducteur (conversion de la pression acoustique en signal électrique) : condensateur

- réponse en fréquence : gamme large et plate (c.-à-d. variation de <2 dB), idéalement 10 Hz-20 000 Hz ; au moins 50 Hz-8000 Hz pour capturer tout le spectre de la parole/voix
- plage dynamique : 15-130 dB ; le niveau de bruit interne du microphone (voir les spécifications du fabricant) doit être inférieur d'au moins 15 dB à la phonation la plus douce, c.-à-d. un niveau de bruit de max. 30 dBA lorsque le microphone est utilisé à une distance de 5 cm de la bouche
- utilisez des embouts en mousse pour éviter les turbulences sur les plosives, les fricatives ou le bruit du flux d'air
- 4–10 cm de la bouche du patient à un angle de 45-90° (essayez de toujours utiliser la même distance, car cela a un impact important sur les mesures d'intensité) ; la distance optimale pour une réponse en fréquence plate (nécessaire pour les mesures spectrales) devrait être indiquée dans les spécifications du microphone

Préamplificateur :

Le préamplificateur amplifie l'amplitude/le voltage faible du microphone :

- utilisez un préamplificateur qui a la même réponse en fréquence et la même plage dynamique que le microphone
- vérifiez l'impédance de terminaison minimale (« minimum terminating impedance ») dans les spécifications du microphone
- le microphone à condensateur doit être alimenté par l'intermédiaire de l'« alimentation fantôme » (« phantom power supply »), généralement 48 V
- niveau d'entrée : réglez-le de manière à ce que les niveaux vocaux maximums soient proches mais inférieurs aux niveaux maximums enregistrables du dispositif (évitez la saturation)

Enregistrement numérique :

- fréquence d'échantillonnage minimale : 44,1 kHz
- résolution minimale en amplitude : 16 bits
- format de fichier : fichier .wav (pas de compression ; ne PAS utiliser .mp3)
- mode d'enregistrement : mono

Références

- [Abraham 20] Abraham, A. K., Pushpavathi, M., Sreedevi, N., Navya, A., Vikram, C. M. M., Prasanna, S. R. M., & Mahadeva Prasannâ, S. R. (2020). Spectral moment and duration of burst of plosives in speech of children with hearing impairment and typically developing children: A comparative study. *Interspeech*, 4981–4985. <https://doi.org/10.21437/Interspeech.2020-1805>
- [Acher 14] Acher, A., Perrier, P., Savariaux, C., & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics and Phonetics*, 28(4), 241–256. <https://doi.org/10.3109/02699206.2013.802015>
- [Barsties 15] Barsties, B., & Maryn, Y. (2015). The improvement of internal consistency of the Acoustic Voice Quality Index. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, 36(5), 647–656. <https://doi.org/10.1016/j.amjoto.2015.04.012>
- [Behrens 88] Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *Journal of Phonetics*, 16(3), 295–298. [https://doi.org/10.1016/s0095-4470\(19\)30504-2](https://doi.org/10.1016/s0095-4470(19)30504-2)
- [Blacklock 04] Blacklock, O. (2004). *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. University of Southampton.
- [Bond 94] Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, 14(4), 325–337. [https://doi.org/10.1016/0167-6393\(94\)90026-4](https://doi.org/10.1016/0167-6393(94)90026-4)
- [Cattelain 19] Cattelain, T. (2019). *Production des consonnes plosives du français : du contrôle des bruits de plosion*. Université Grenoble Alpes.
- [Chial 03] Chial, M. R. (2003). Suggestions for computer-based audio recording of speech samples for perceptual and acoustic analyses. *Phonology Project Technical Report No. 13*.
- [Delgado-Hernandez 17] Delgado, J., León, N., Jiménez, A., & Izquierdo, L. (2017). Acoustic analysis of the voice: Temporal, spectral and cepstral measurements in the normal voice with the Praat in a sample of Spanish speakers. *Revista de Investigacion en Logopedia*, 7(2), 108-127.
- [Delgado-Hernandez 18] Delgado-Hernández, J., León-Gómez, N. M., Izquierdo-Arteaga, L. M., & Llanos-Fumero, Y. (2018). Análisis cepstral de la voz normal y patológica en adultos españoles. Medida de la prominencia del pico cepstral suavizado en vocales sostenidas versus habla conectada. *Acta Otorrinolaringológica Española*, 69(3), 134–140. <https://doi.org/10.1016/j.otorri.2017.05.006>
- [Deliyski 05] Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19(1), 15–28. <https://doi.org/10.1016/j.jvoice.2004.07.003>
- [Dromey 03] Dromey, C. (2003). Spectral measures and perceptual ratings of hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology*, 11(2), 85–94. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0038362549&partnerID=tZOTx3y1>
- [Gottsmann 11] Gottsmann, F., & Harwardt, C. (2011). Investigating robustness of spectral moments on normal-and high-effort speech. *Interspeech, August*, 2937–2940.
- [Hammarberg 80] Hammarberg, B., Fritzen, B., Gauffin, J., Sundberg, J. & Wedin, L. (1980) Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, 90, 441- 451.
- [Heman-Ackah 03] Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., Horman, M., & Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology*, 112(4), 324–333. <https://doi.org/10.1177/000348940311200406>
- [Hernandez 19] Hernandez, A., Lee, H., & Chung, M. (2019). Acoustic analysis of fricatives in dysarthric speakers with cerebral palsy. *Phonetics and Speech Sciences*, 11(3), 23–29. <https://doi.org/10.13064/ksss.2019.11.3.023>
- [Hillenbrand 94] Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778. <https://doi.org/10.1044/jshr.3704.769>
- [Hillenbrand 96] Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39, 311-321. <https://doi.org/10.1044/jshr.3902.311>
- [Iskarous 11] Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2), 944–954. <https://doi.org/10.1121/1.3514537>
- [Jongman 00] Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252. <https://doi.org/10.1121/1.1288413>

- [Kay 12] Kay, T. S. (2012). *Spectral analysis of stop consonants in individuals with dysarthria secondary to stroke*. Louisiana State University.
- [Kent 89] Kent, R. D., Kent, J. F., Weismer, G., Martin, R. E., Sufit, R. L., Brooks, B. R., & Rosenbek, J. C. (1989a). Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects. *Clinical Linguistics & Phonetics*, 3(4), 347–358. <https://doi.org/10.3109/02699208908985295>
- [Kent 99] Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*, 32(3), 141–186. [https://doi.org/10.1016/S0021-9924\(99\)00004-0](https://doi.org/10.1016/S0021-9924(99)00004-0)
- [Y. Kim 09] Kim, Y., Weismer, G., Kent, R. D., & Duffy, J. R. (2009). Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatrica et Logopaedica*, 61(6), 329–335. <https://doi.org/10.1159/000252849>
- [Körkkö 15] Körkkö, P. (2015). Spectral moments analysis of /s/ coarticulation development in Finnish-speaking children. *ICPhS*. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0470.pdf>
- [Kuruville-Dugdale 20] Kuruville-Dugdale, M., Dietrich, M., McKinley, J. D., & Deroche, C. (2020). An exploratory model of speech intelligibility for healthy aging based on phonatory and articulatory measures. *Journal of Communication Disorders*, 87(March), 105995. <https://doi.org/10.1016/j.jcomdis.2020.105995>
- [Lam 16] Lam, J., & Tjaden, K. (2016). Clear speech variants: An acoustic study in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 59(4), 631–646. https://doi.org/10.1044/2015_JSLHR-S-15-0216
- [Liu 05] Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117(6), 3879–3889. <https://doi.org/10.1121/1.1898623>
- [Lowell 11] Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25(5), e223–e232. <https://doi.org/10.1016/j.jvoice.2010.06.007>
- [Master 06] Master, S., De Biase, N., Pedrosa, V., & Chiari, B. M. (2006). The long-term average spectrum in research and in the clinical practice of speech therapists. *Pró-Fono Revista de Atualização Científica*, 18(1), 111–120. <http://doi.org/10.1590/S0104-56872006000100013>
- [Meyer 11] Meyer, S., & Ptok, M. (2011). Zur objektiven Beurteilung der Zischlautbildung. *Laryngo-Rhino-Otologie*, 90(12), 753–759. <https://doi.org/10.1055/s-0031-1275279>
- [Munson 04] Munson, B. (2004). Children and adults: Evidence spectral mean. *Journal of Speech, Language, and Hearing Research*, 47(February), 58–69.
- [Newman 03] Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5), 2850–2860. <https://doi.org/10.1121/1.1567280>
- [Nissen 03] Nissen, S. (2003). An acoustic analysis of voiceless obstruents produced by adults and typically developing children. Ohio State University. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1041225568&disposition=inline
- [Nissen 05] Nissen, S., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118(4), 2570–2578. <https://doi.org/10.1121/1.2010407>
- [Patel 18] Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyiski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009
- [Plichta 02] Plichta, B. (2002). Best practices in the acquisition, processing, and analysis of acoustic speech signals. *U. Penn Working Papers in Linguistics*, 8(3), 209–222. <http://flint.matrix.msu.edu/extras/Audio-technology.pdf%0Ahttp://repository.upenn.edu/pwpl/vol8/iss3/16%0Ahttp://repository.upenn.edu/pwpl/vol8/>
- [Repp 89] Repp, B. H., & Lin, H. B. (1989). Acoustic properties and perception of stop consonant release transients. *Journal of the Acoustical Society of America*, 85(1), 379–396. <https://doi.org/10.1121/1.397689>

- [Sapir 09] Sapir, S., Polczyńska, M., & Tobin, Y. (2009). Why does the vowel space area as an acoustic metric fail to differentiate dysarthric from normal vowel articulation and what can be done about it? *Poznan Studies in Contemporary Linguistics*, 45(2), 301–311. <https://doi.org/10.2478/v10010-009-0018-2>
- [Sauder 17] Sauder, C., Bretl, M., & Eadie, T. (2017). Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice*, 31(5), 557–566. <https://doi.org/10.1016/j.jvoice.2017.01.006>
- [Scarbel 12] Scarbel, L. (2012). *Etude acoustique de la production de consonnes et de voyelles du français par des enfants porteurs d'un implant cochléaire*. Université Grenoble Alpes.
- [Smith 14] Smith, L. K., & Goberman, A. M. (2014). Long-time average spectrum in individuals with Parkinson disease. *NeuroRehabilitation*, 35(1), 77–88. <https://doi.org/10.3233/NRE-141102>
- [Šrámková 15] Šrámková, H., Granqvist, S., Herbst, C. T., & Švec, J. G. (2015). The softest sound levels of the human voice in normal subjects. *The Journal of the Acoustical Society of America*, 137(1), 407–418. <https://doi.org/10.1121/1.4904538>
- [Stevens 98] Stevens, K. N. (1998). *Acoustic Phonetics*. MIT Press.
- [Švec 10] Švec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, 19(4), 356–368. [https://doi.org/10.1044/1058-0360\(2010/09-0091\)](https://doi.org/10.1044/1058-0360(2010/09-0091))
- [Švec 18] Švec, J. G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research*, 61(3), 441–461. https://doi.org/10.1044/2017_JSLHR-S-17-0095
- [Tjaden 97] Tjaden, K., & Turner, G. S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, 40(6), 1358–1372. <https://doi.org/10.1044/jslhr.4006.1358>
- [Tjaden 10] Tjaden, K., Sussman, J. E., Liu, G., & Wilding, G. (2010). Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. *Journal of Medical Speech-Language Pathology*, 18(4), 125–132. <https://doi.org/10.1016/j.artmed.2015.09.007.Information>
- [Tjaden 17] Tjaden, K., & Martel-Sauvageau, V. (2017). Consonant acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions. *American Journal of Speech-Language Pathology*, 26(2Special Issue), 569–582. https://doi.org/10.1044/2017_AJSLP-16-0090
- [Van Son 18] Van Son, R. J. J. H., Middag, C., & Demuyne, K. (2018). Vowel space as a tool to evaluate articulation problems. *Interspeech*, 357–361.
- [Volin 11] Volin, J., & Zimmermann, J. (2011). Spectral slope parameters and detection of word stress. *Proceedings of the Technical Computing Prague*.
- [Weismer 01] Weismer, G., Jeng, J.-Y., Laues, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica*, 53(1), 1–18. <https://doi.org/10.1159/000052649>
- [Weismer 92] Weismer, G., Martin, R., Kent, R. D., & Kent, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *The Journal of the Acoustical Society of America*, 91(2), 1085–1098. <https://doi.org/10.1121/1.402635>
- [Weismer 95] Weismer, G., Tjaden, K., & Kent, R. D. (1995). Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? *Journal of Phonetics*, 23(1–2), 149–164. [https://doi.org/10.1016/S0095-4470\(95\)80039-5](https://doi.org/10.1016/S0095-4470(95)80039-5)
- [Winholtz 97] Winholtz, W. S., & Titze, I. R. (1997). Miniature head-mounted microphone for voice perturbation analysis. *Journal of Speech, Language, and Hearing Research*, 40(4), 894–899. <https://doi.org/10.1044/jslhr.4004.894>
- [Yoon 15] Yoon, T.-J. (2015). A Corpus-based study on the effects of gender on voiceless fricatives in American English. *Phonetics and Speech Sciences*, 7(1), 117–124. <https://doi.org/10.13064/ksss.2015.7.1.117>