CNRS - INP - UT3 - UT1 - UT2J
Institut de Recherche en Informatique de Toulouse

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

Hôpitaux de Toulouse

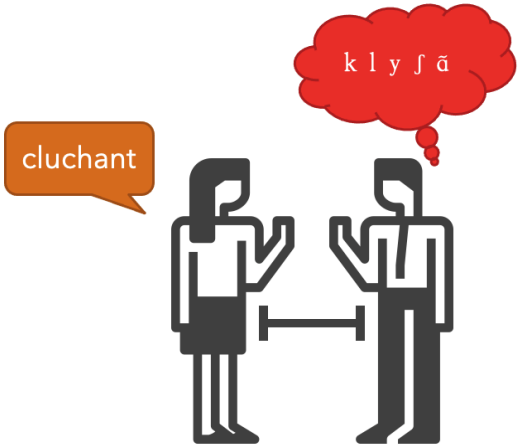
Mesure de l'intelligibilité après cancer oral ou oropharyngé par un système de reconnaissance automatique de la parole

Mathieu Balaguer,
Lucile Gelin, Virginie Woisard, Jérôme Farinas, Julien Pinquier

Contexte

- Cancers de la cavité buccale et de l'oropharynx (Santé publique France, 2019)
  - Fréquents : 13 692 cas en 2018
  - Localisation lèvres-bouche-pharynx : conséquences sur les capacités de parole (Barrett et al., 2004; Borggreven et al., 2007; Colangelo et al., 2000; DeNittis et al., 2001; Mlynarek et al., 2008; Stelzle et al., 2013)

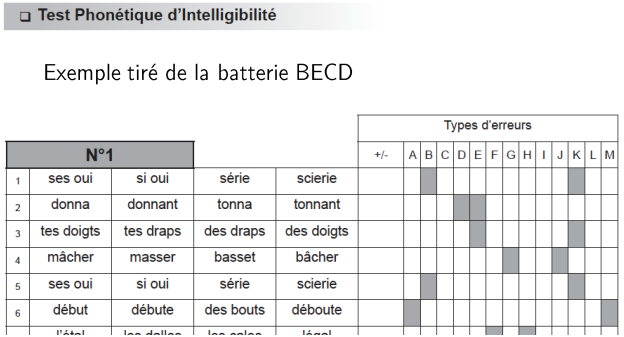
- Dégradation de l'intelligibilité de la parole
  - Concerne les déficits de bas niveau (Lindblom, 1990; Hustad, 2008)
  - Définition consensuelle après enquête DELPHI (Pommée et al., 2021)
    - Relative à la reconstruction d'un énoncé au niveau acoustico-phonétique
    - Informations relatives à l'intelligibilité véhiculées par le signal acoustique
    - En jeu dans les énoncés où les processus cognitifs de compensation ne sont pas mis en jeu : pseudo-mots, phrases non prévisibles…



Contexte

- Évaluation perceptive de l'intelligibilité
  - Évaluation globale au moyen d'échelles de Likert
    - Intelligibilité globale
    - Paramètres spécifiques type « distorsions phonétiques »

  - Reconnaissance de pseudo-mots
    - Transcriptions de syllabes ou pseudo-mots



Exemple tiré de la batterie BECD

N°1					Types d'erreurs												
					+/-	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ses oui	si oui	série	scierie			■									■		
2	donna	donnant	tonna	tonnant					■	■								
3	tes doigts	tes draps	des draps	des doigts						■						■		
4	mâcher	masser	basset	bâcher							■				■			
5	ses oui	si oui	série	scierie			■									■		
6	début	débute	des bouts	déboute		■						■						■

  - Limites
    - Manque de standardisation des épreuves (Pommée et al., 2020) et des outils de mesure
    - Variabilités inter- et intra-juges (Fex, 1992; Van Nuffelen et al., 2009)

Contexte

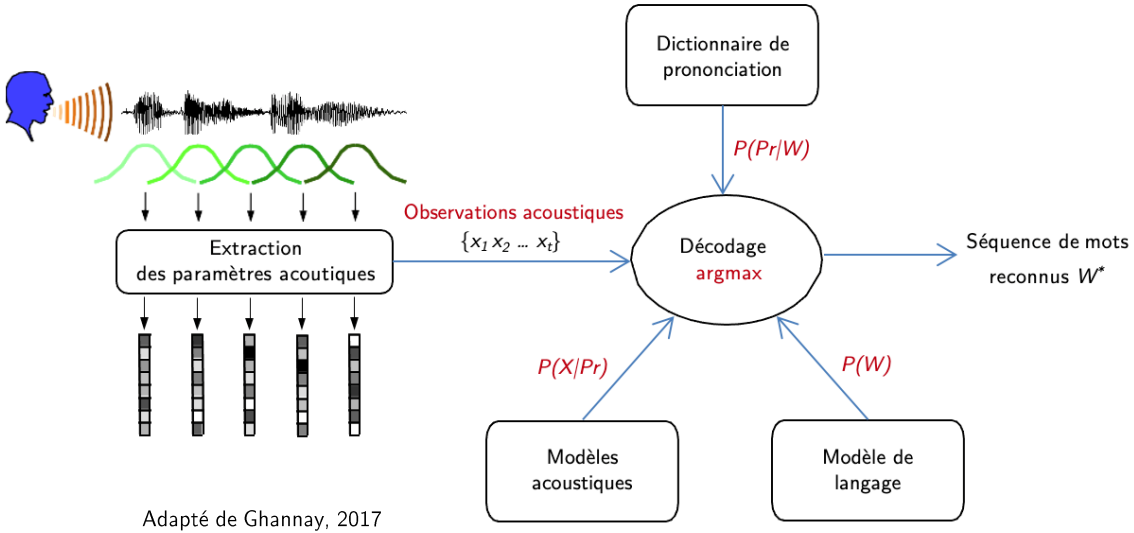
- Évaluation automatique

  - Développement récent d'outils utilisables en pratique courante
    - MonPaGe (https://lpp.in2p3.fr/monpage/; Pernon et al., 2020)
    - Traitement automatique de tâches de répétition de pseudo-mots (projets C2SI, RUGBI)

  - Quelle tâche support de parole ? (Balaguer et al., 2020)
    - Analyses acoustiques et automatiques portent majoritairement
      - Sur des phonèmes isolés
      - Plus spécifiquement, sur des voyelles tenues, voire sur des phonèmes extraits de texte
    - Pas d'étude menée sur de véritables situations de parole spontanée non contrainte
      - Pourtant la plus représentative de l'expression orale dans la vie quotidienne (Knuijt et al., 2017)

Contexte

- Utilisation de systèmes de reconnaissance automatique de la parole (RAP)
  - Permettent de produire une séquence d'unités (mots ou sons élémentaires = phones) à partir du signal acoustique selon une approche probabiliste

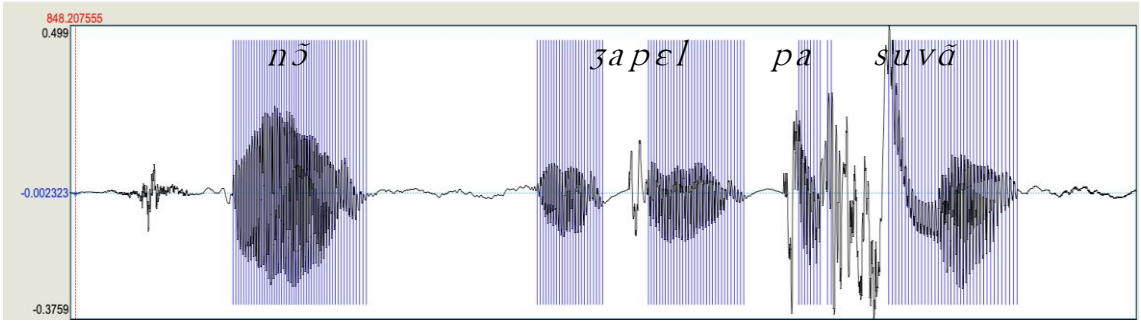


Adapté de Ghannay, 2017

Objectif

Objectif

Prédire l'intelligibilité de la parole après traitement d'un cancer oral ou oropharyngé au moyen de scores issus d'un système de reconnaissance automatique de la parole



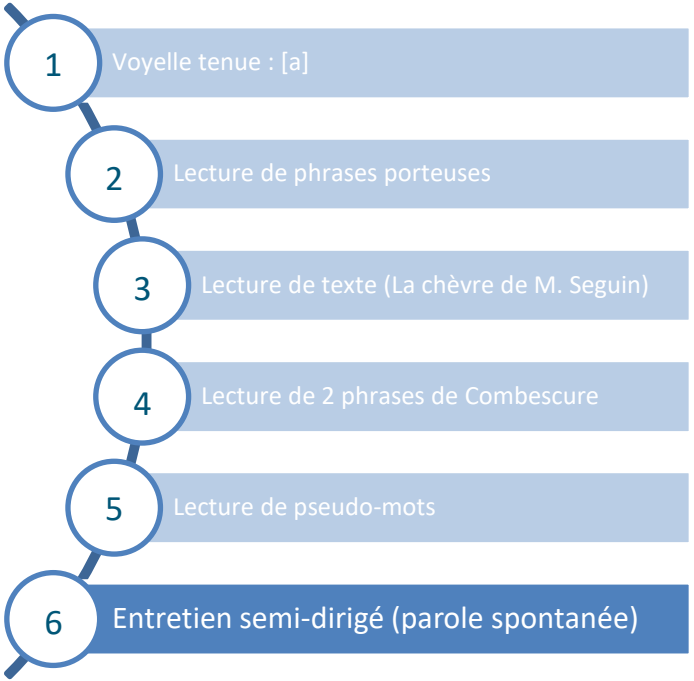
Inclusion / Non inclusion

- Projet RUGBI (ANR-18-CE45-0008)

Critères d'inclusion	Critères de non inclusion
▪ Adultes ▪ Francophones natifs ▪ Traités pour un cancer de la cavité buccale ou de l'oropharynx depuis plus de 6 mois ▪ En rémission clinique depuis plus de 6 mois	▪ Patients fatigables ▪ Autre pathologie responsable d'un trouble de la parole ▪ Présence de troubles cognitifs

Enregistrements de parole

- Enregistrements de parole spontanée au cours d'un entretien semi-dirigé
  - Contexte d'entretien semi-dirigé équivaut à une véritable situation de parole spontanée (Prins & Bastiaanse, 2004)
  - Salle non anéchoïque
    - Enregistreur numérique
    - Micro serre-tête 6 cm de la bouche du sujet



Analyse perceptive

- Analyse perceptive de la parole

  – Jugement perceptif (mesure de référence de la sévérité du trouble de parole) (Balaguer et al., 2019)
    - 6 auditeurs « experts »
    - Évaluation sur une échelle de 0 (très sévère) à 10 (absence de trouble) :

https://care.easiware.fr/fr-FR/Post/56

  – Intelligibilité (de 0 – *très sévère* – à 10 – *absence de trouble*)

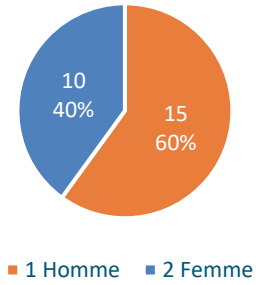
Analyse automatique de la parole

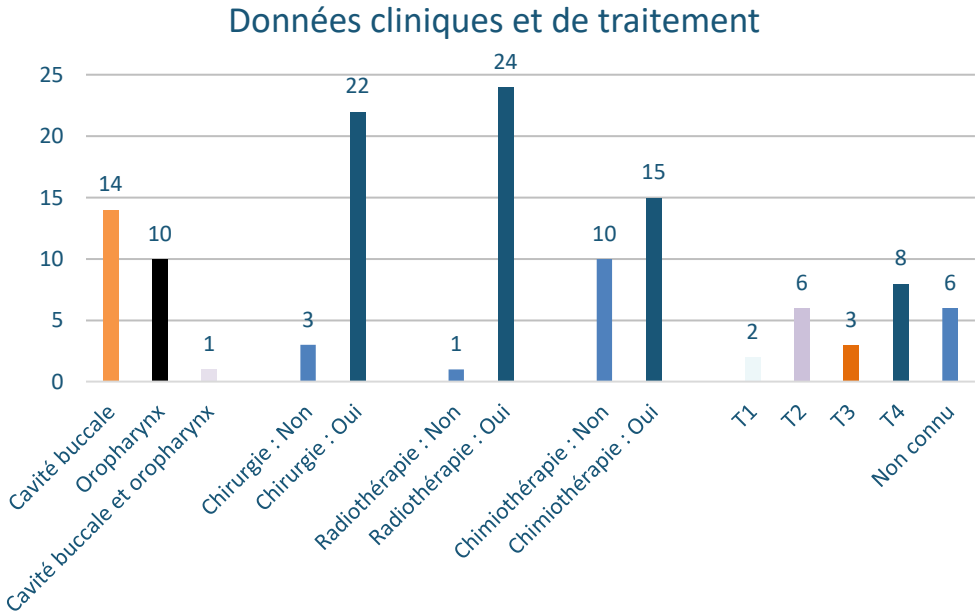
- Analyse automatique de la parole (RAP)
  - Architecture de type Time-Delay Neural Network factorisé ou TDNNf-HMM (Povey, 2018)
  - Adapté à la parole non typique (Gelin et al., 2021)

Entraînement	Décodage
▪ Corpus Common Voice, septembre 2019 (parole typique) : https://commonvoice.mozilla.org/fr	▪ Segments de parole du sujet seul, sur la tâche d'entretien semi-dirigé (Google WebRTC-VAD https://github.com/wiseman/py-webrtcvad) ▪ Détection parmi 33 phonèmes du français 18 consonnes, 12 voyelles, 3 semi-consonnes ▪ Score de confiance associé à chaque phonème par méthode Minimum Bayes Risk (Xu et al., 2011)

Résultats

- Description de l'échantillon



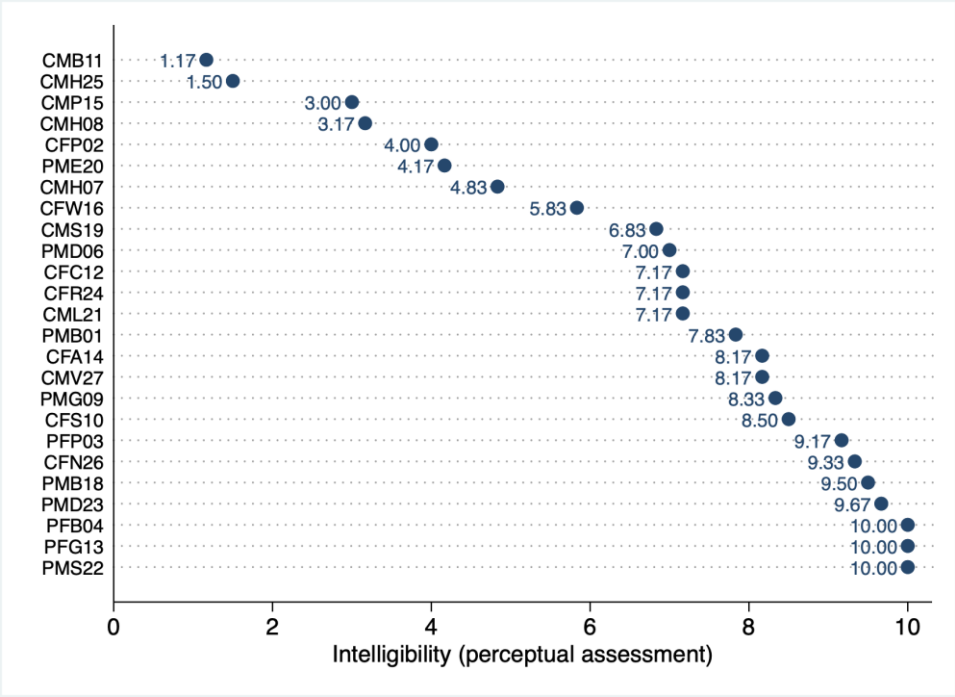


Délai depuis traitement

m = 87 mois (Me = 40)
s = 121 mois (EIQ = 123)

Analyse perceptive

- Évaluation perceptive de l'intelligibilité



Analyse automatique

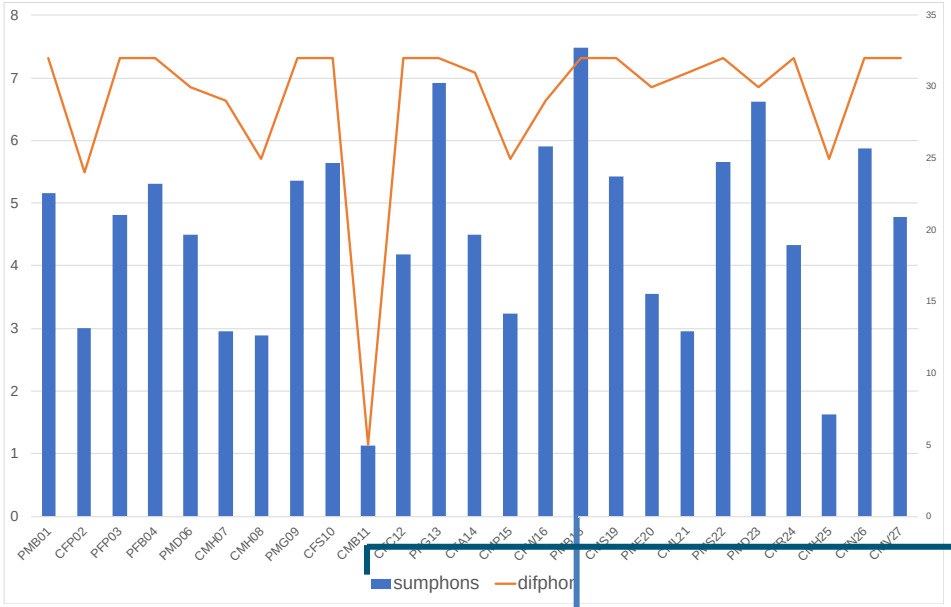
- Inventaire phonémique, par sujet
  - Nombre de phonèmes reconnu par seconde (1 paramètre)
  - Nombre total de phonèmes différents reconnus (1 paramètre)
  - Plus précisément
    - Nombre de consonnes, de voyelles, de semi-consonnes, d'occlusives, de fricatives, de sonantes, de non sonantes par seconde (7 paramètres)
    - Proportion de chaque type (7 paramètres)

PFB04-ECV-16k_mono_299_84999999996904_303_26999999996593	M E N a A~ e P A E E a M D F a M V P a T o~ T R E~
PFB04-ECV-16k_mono_304_4999999999648_305_2799999999641	
PFB04-ECV-16k_mono_305_8199999999636_308_54999999996113	O E M E E~ K u R S O M Z H E N P E~ B G e E~
PFB04-ECV-16k_mono_313_6199999999565_314_6399999999556	M y o~
PFB04-ECV-16k_mono_321_9899999999489_322_6499999999483	
PFB04-ECV-16k_mono_322_8899999999481_324_14999999994694	N o~
PFB04-ECV-16k_mono_324_2099999999469_325_2899999999459	
PFB04-ECV-16k_mono_328_6199999999429_329_9099999999417	L Z H E N
PFB04-ECV-16k_mono_32_610000000000475_35_28000000000058	M W E~ D F W a O E Z
PFB04-ECV-16k_mono_330_29999999994135_331_61999999994015	
PFB04-ECV-16k_mono_332_9999999999389_334_5599999999375	Z H E R e a R Z H
PFB04-ECV-16k_mono_336_11999999993606_338_0699999999343	a o~
PFB04-ECV-16k_mono_339_5699999999329_340_61999999993196	B e A~
PFB04-ECV-16k_mono_346_55999999992656_348_20999999992506	Z H a G u F N

Analyse automatique

JSO2021

Phonème	Nombre d'occurrences
a	401
E	259
S	244
L	230
R	220
P	200
K	189
T	170
i	168
M	156
A~	151
E~	128
D	118
e	114
N	95
ZH	91
y	88
O	88
AE	86
V	82
F	68
J	65
u	51
B	45
o~	38
SH	28
Z	23
W	20
G	17
o	16
H	4
NJ	2



Phonème	Nombre d'occurrences
a	110
R	21
E~	8
A~	1
AE	1

Analyse automatique

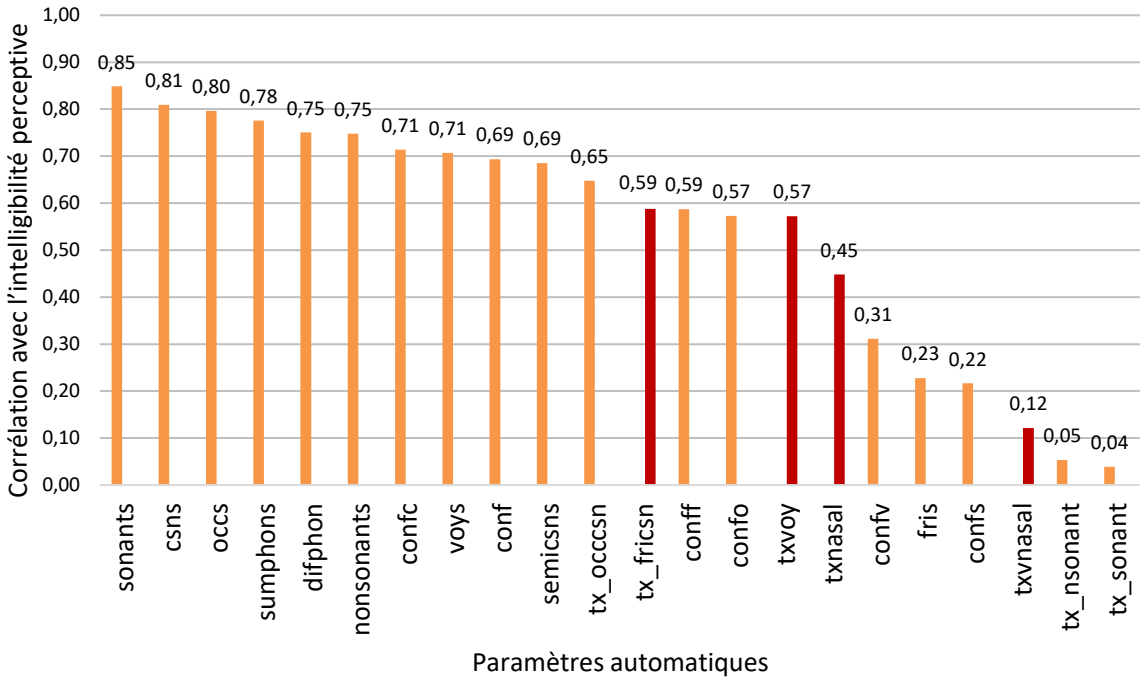
- Analyses des scores de confiance à la sortie du TDNN
  - Calcul
    - Score de confiance moyen global
      (1 paramètre)
    - Score de confiance moyen sur les consonnes, les voyelles, les semi-consonnes, les occlusives et les fricatives
      (5 paramètres)

CFC12-ECV-16k_mono_1012_9799999993205_1017_929999999316	1	4.30	0.23	AE	0.76
CFC12-ECV-16k_mono_1018_2899999993157_1018_9499999993151	1	0.03	0.03	P	0.94
CFC12-ECV-16k_mono_1018_2899999993157_1018_9499999993151	1	0.06	0.16	a	1.00
CFC12-ECV-16k_mono_1018_2899999993157_1018_9499999993151	1	0.22	0.08	R	0.71
CFC12-ECV-16k_mono_1023_1199999993113_1036_829999999299	1	0.03	0.10	J	0.48
CFC12-ECV-16k_mono_1023_1199999993113_1036_829999999299	1	0.81	0.06	V	0.38

CFC12	
Score global	0,82773995
Consonnes	0,86108392
Voyelles	0,79563898
Semi-consonnes	0,72224998
Occlusives	0,92210525
Fricatives	0,84257793

Modélisation

- Analyse bivariée entre chacun des 22 paramètres et l'intelligibilité
  - Coefficients de corrélation de Spearman



Résultats

- Modélisation par sélection des paramètres au moyen d'une régression LASSO (Tibshirani, 1996)
  - Imputation des deux données manquantes : scores de confiance occlusives et semi-consonnes de CMB11
  - 3 variables explicatives retenues :
    - **proportion d'occlusives parmi les consonnes (tx_occcsn)**
    - **nombre de sonantes par seconde (sonants)**
    - **score de confiance moyen sur les fricatives (conff)**
    - ~~**nombre d'occlusives par seconde (occs)**~~ *(multicolinéarité)*

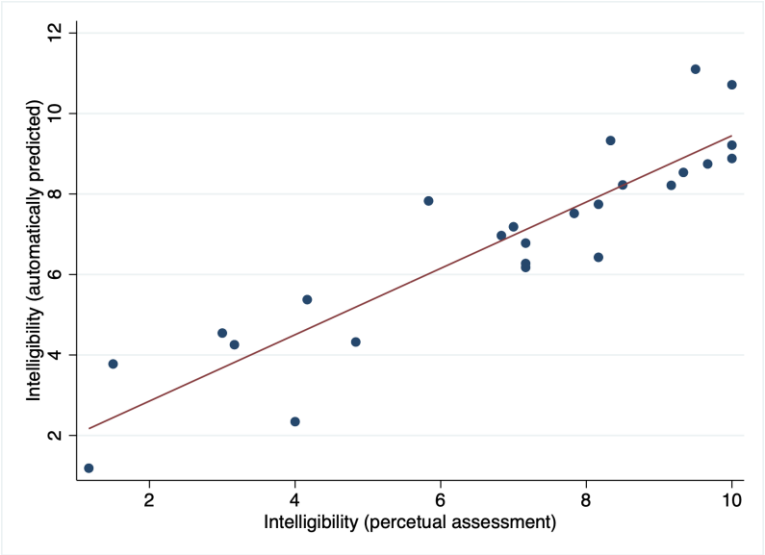
$$Intelligibilité = -0{,}073 + (6{,}188 \times \boldsymbol{tx\_occcsn}) + (4{,}982 \times \boldsymbol{sonants}) + (0{,}851 \times \boldsymbol{conff})$$

Résultats

- Modélisation par sélection des paramètres au moyen d'une régression LASSO
  (Tibshirani, 1996)

$$Intelligibilité = -0{,}073 + (6{,}188 \times \boldsymbol{tx\_occcsn}) + (4{,}982 \times \boldsymbol{sonants}) + (0{,}851 \times \boldsymbol{conff})$$

- RMSE=1,21
- R^2=0,824
- Corrélation entre intelligibilité perceptive et intelligibilité prédite : r_S=0,91 (p<0,001)
- Validation croisée :
  - 5 blocs : r_S=0,90 (p<0,001)
  - « Leave one out » : r_S=0,90 (p<0,001)



Conclusion

L'intelligibilité peut être prédite de façon correcte et fiable au moyen de 3 paramètres issus d'une analyse automatique de la parole spontanée par un système de RAP

Élargir la taille de l'échantillon de l'étude

Optimiser la mesure automatique de la parole pathologique

Rendre les résultats applicables en clinique courante

- Généralisabilité des résultats
- Mise en évidence de déficits plus fins après entraînement sur de la parole pathologique ?
  - Manque de corpus +++
  - Entraînement sur de la parole pathologique par transfer learning (Gelin et al., 2021; Wang et al., 2015)
- Intérêts cliniques des systèmes de reconnaissance automatique de la parole
  - Applicables à l'étude de la parole spontanée
  - Scores fiables
  - Équipement requis minimal et peu onéreux
- Développement d'un dispositif mobile
  - Expérimentation en cours au CHU de Toulouse concernant la parole

CNRS - INP - UT3 - UT1 - UT2J
Institut de Recherche en Informatique de Toulouse

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

Hôpitaux de Toulouse

Mesure de l'intelligibilité après cancer oral ou oropharyngé par un système de reconnaissance automatique de la parole

Mathieu Balaguer,
Lucile Gelin, Virginie Woisard, Jérôme Farinas, Julien Pinquier

Références

Balaguer, M., Boisguérin, A., Galtier, A., Gaillard, N., Puech, M., & Woisard, V. (2019). Assessment of impairment of intelligibility and of speech signal after oral cavity and oropharynx cancer. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *136*(5), 355–359. https://doi.org/10.1016/j.anorl.2019.05.012

Balaguer, M., Pommée, T., Farinas, J., Pinquier, J., Woisard, V., & Speyer, R. (2020). Effects of oral and oropharyngeal cancer on speech intelligibility using acoustic analysis: Systematic review. *Head & Neck*, *42*(1), 111–130. https://doi.org/10.1002/hed.25949

Barrett, W. L., Gluckman, J. L., Wilson, K. M., & Gleich, L. L. (2004). A comparison of treatments of squamous cell carcinoma of the base of tongue: surgical resection combined with external radiation therapy, external radiation therapy alone, and external radiation therapy combined with interstitial radiation. *Brachytherapy*, *3*(4), 240–245. https://doi.org/10.1016/j.brachy.2004.09.002

Borggreven, P. A., Verdonck-De Leeuw, I. M., Muller, M. J., Heiligers, M. L. C. H., De Bree, R., Aaronson, N. K., & Leemans, C. R. (2007). Quality of life and functional status in patients with cancer of the oral cavity and oropharynx: Pretreatment values of a prospective study. *European Archives of Oto-Rhino-Laryngology*, *264*(6), 651–657. https://doi.org/10.1007/s00405-007-0249-5

Colangelo, L. A., Logemann, J. A., & Rademaker, A. W. (2000). Tumor Size and Pretreatment Speech and Swallowing in Patients with Resectable Tumors. *Otolaryngology–Head and Neck Surgery*, *122*(5), 653–661. https://doi.org/10.1016/S0194-5998(00)70191-4

DeNittis, A. S., Machtay, M., Rosenthal, D. I., Sanfilippo, N. J., Lee, J. H., Goldfeder, S., Chalian, A. A., Weinstein, G. S., & Weber, R. S. (2001). Advanced oropharyngeal carcinoma treated with surgery and radiotherapy: Oncologic outcome and functional assessment. *American Journal of Otolaryngology*, *22*(5), 329–335. https://doi.org/10.1053/ajot.2001.26492

Fex, S. (1992). Perceptual evaluation. *Journal of Voice*, *6*(2), 155–158.

Gelin, L., Daniel, M., Pinquier, J., & Pellegrini, T. (2021). *End-to-end acoustic modelling for phone recognition of young readers*. https://www.lalilo.com/

Hustad, K. C. (2008). The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria. *Journal of Speech, Language, and Hearing Research*, *51*(3), 562–573. https://doi.org/10.1044/1092-4388(2008/040)

Knuijt, S., Kalf, J. G., van Engelen, B. G. M., de Swart, B. J. M., & Geurts, A. C. H. (2017). The Radboud Dysarthria Assessment: Development and Clinimetric Evaluation. *Folia Phoniatrica et Logopaedica*, *69*(4), 143–153. https://doi.org/10.1159/000484556

Lindblom, B. (1990). On the Communication Process: Speaker-Listener Interaction and the Development of Speech. *Augmentative and Alternative Communication*, *6*(4), 220–230. https://doi.org/10.1080/07434619012331275504

Middag, C., Martens, J. P., Van Nuffelen, G., & De Bodt, M. (2009). Automated Intelligibility Assessment of Pathological Speech Using Phonological Features. *Eurasip Journal on Advances in Signal Processing*, *2009*. https://doi.org/10.1155/2009/629030

Références

Mlynarek, A., Rieger, J., Harris, J., O'Connell, D., Al-Qahtani, K., Ansari, K., Chau, J., & Seikaly, H. (2008). Methods of functional outcomes assessment following treatment of oral and oropharyngeal cancer: review of the literature. *Journal of Otolaryngology - Head & Neck Surgery*, *37*(1), 2–10. https://doi.org/10.2310/7070.2008.1001

Pernon, M., Lévêque, N., Delvaux, V., Assal, F., Borel, S., Fougeron, C., Trouville, R., & Laganaro, M. (2020). MonPaGe, un outil de screening francophone informatisé d'évaluation perceptive et acoustique des troubles moteurs de la parole (dysarthries, apraxie de la parole. *Rééducation Orthophonique*, *281*(January), 169–198.

Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021a). Assessment of adult speech disorders: current situation and needs in French-speaking clinical practice. *Logopedics Phoniatrics Vocology*, *0*(0), 1–15. https://doi.org/10.1080/14015439.2020.1870245Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021b). Intelligibility and comprehensibility: A Delphi consensus study. *International Journal of Language & Communication Disorders*, 1–44. https://doi.org/10.1111/1460-6984.12672

Povey, D., Cheng, G., Wang, Y., Li, K., Xu, H., Yarmohammadi, M., & Khudanpur, S. (2018). Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks. *Interspeech 2018*, *2018-Septe*(2), 3743–3747. https://doi.org/10.21437/Interspeech.2018-1417

Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*(12), 1075–1091. https://doi.org/10.1080/02687030444000534

Stelzle, F., Knipfer, C., Schuster, M., Bocklet, T., Nöth, E., Adler, W., Schempf, L., Vieler, P., Riemann, M., Neukam, F. W., & Nkenke, E. (2013). Factors influencing relative speech intelligibility in patients with oral squamous cell carcinoma: A prospective study using automatic, computer-based speech analysis. *International Journal of Oral and Maxillofacial Surgery*, *42*(11), 1377–1384. https://doi.org/10.1016/j.ijom.2013.05.021

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wang, D., & Zheng, T. F. (2015). Transfer learning for speech and language processing. *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 1225–1237. https://doi.org/10.1109/APSIPA.2015.7415532

Xu, H., Povey, D., Mangu, L., & Zhu, J. (2011). Minimum Bayes Risk decoding and system combination based on a recursion for edit distance. *Computer Speech & Language*, *25*(4), 802–828. https://doi.org/10.1016/j.csl.2011.03.001