



HAL
open science

La laryngophoniatrie du futur

Alain Ghio, Jérôme Farinas

► **To cite this version:**

Alain Ghio, Jérôme Farinas. La laryngophoniatrie du futur. Congrès National de la Société Française d'ORL, 2021, Paris, France. hal-03365266v3

HAL Id: hal-03365266

<https://hal.science/hal-03365266v3>

Submitted on 9 Oct 2021

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

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



127^{ÈME} CONGRÈS · PRÉSENTIEL & VIRTUEL

SFORL



DU 01 AU 03
OCTOBRE 2021



PALAIS DES CONGRÈS DE PARIS
& SFORL.ORG

LA LARYNGOPHONIATRIE DU FUTUR



JÉRÔME FARINAS, INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE,
UNIVERSITE PAUL SABATIER

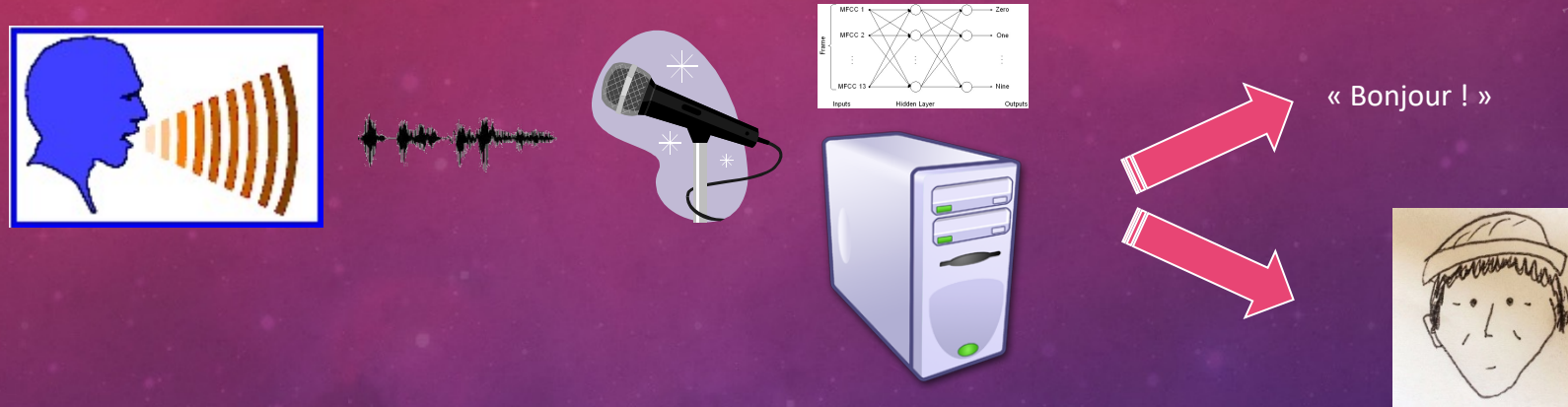


ALAIN GHIO, LABORATOIRE PAROLE ET LANGAGE, CNRS, AIX-MARSEILLE UNIVERSITE

TECHNOLOGIE MONDIALE



TRAITEMENTS AUTOMATIQUES DE LA PAROLE



- Reconnaissance Automatique de la Parole
 - Décodage du contenu audio
 - Adapté pour la mesure de l'intelligibilité
- Reconnaissance automatique du locuteur

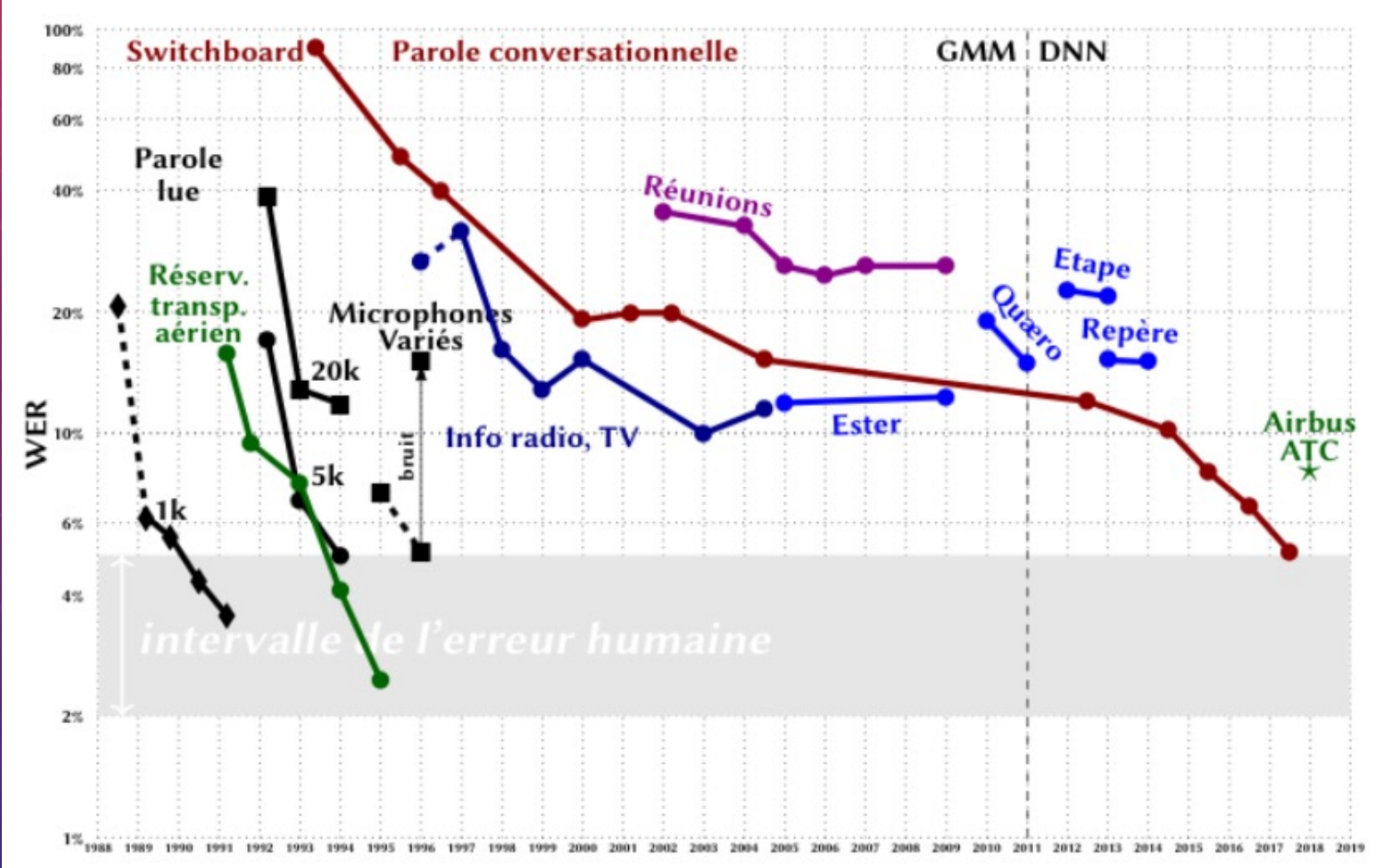
POURQUOI LE TAP?

- Simple
 - Un microphone
 - Traitements automatiques
- Parole spontanée
- Peut analyser des relations complexes entre
 - Entrée (locuteur, parole)
 - Sortie

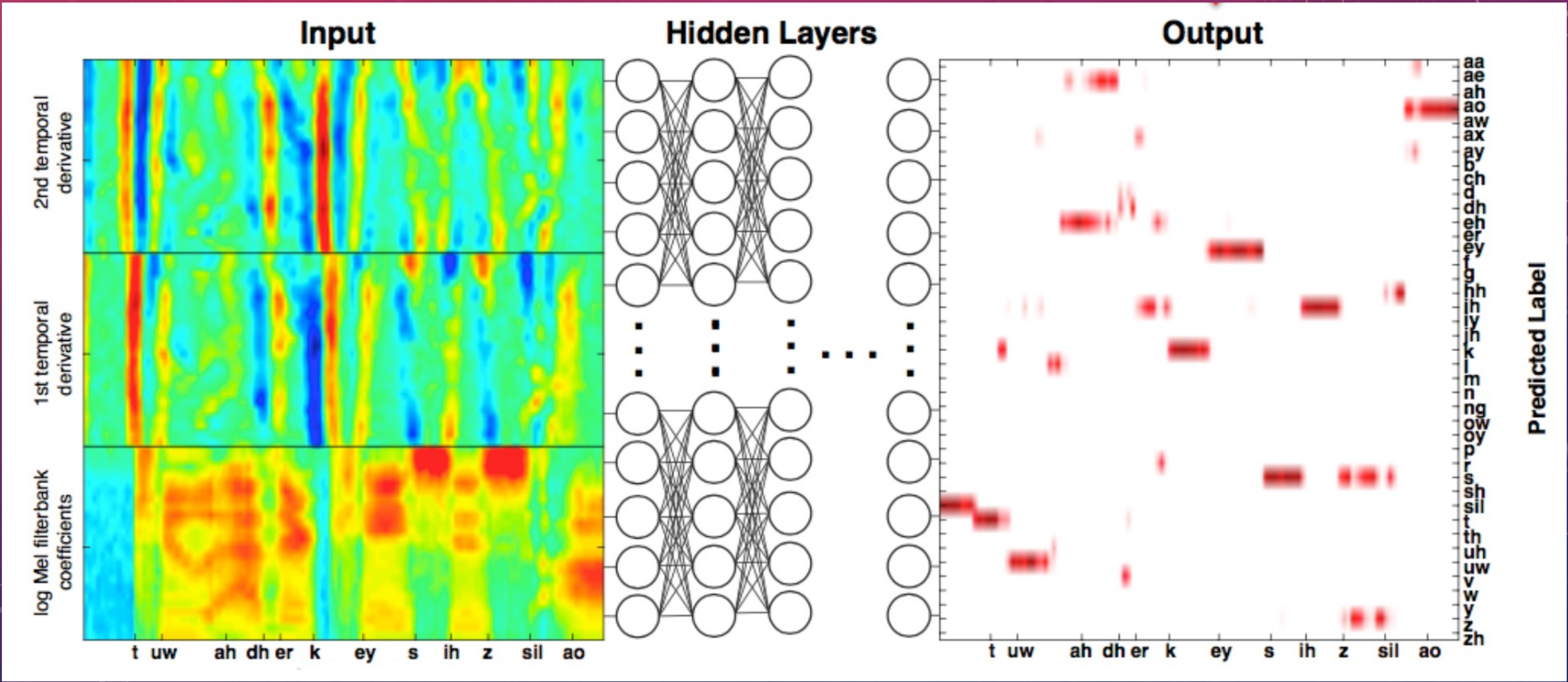
RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

- Recherche à transcrire en texte un enregistrement audio
- Principe
 - Extraction de caractéristiques à partir d'un enregistrement audio
 - Décodage acoustico-phonétique
 - Modélisation du langage

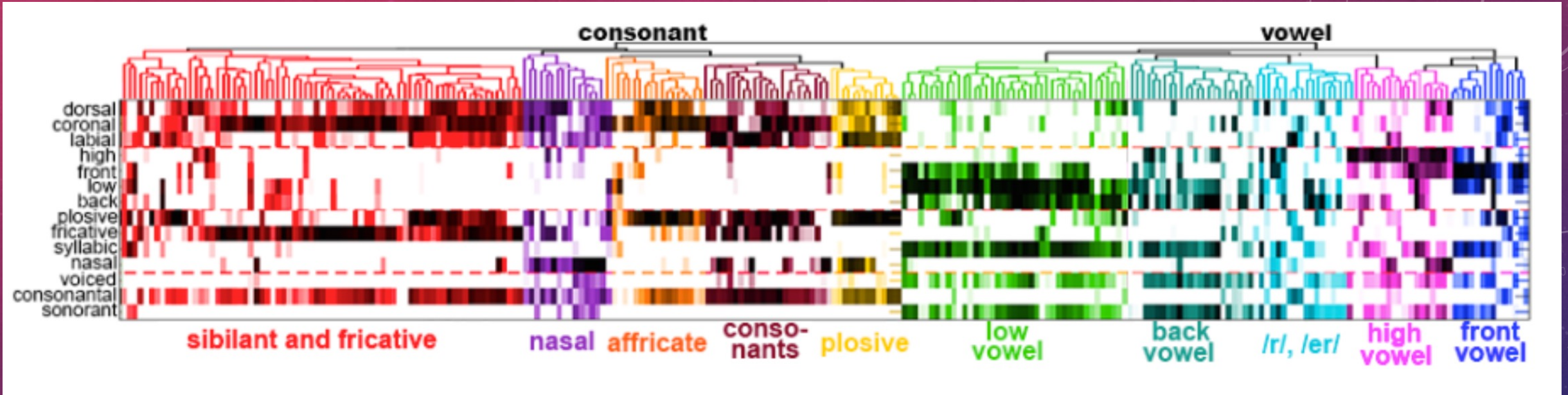
QUOI DE NEUF EN RAP ?



APPRENTISSAGE PROFOND EN RAP



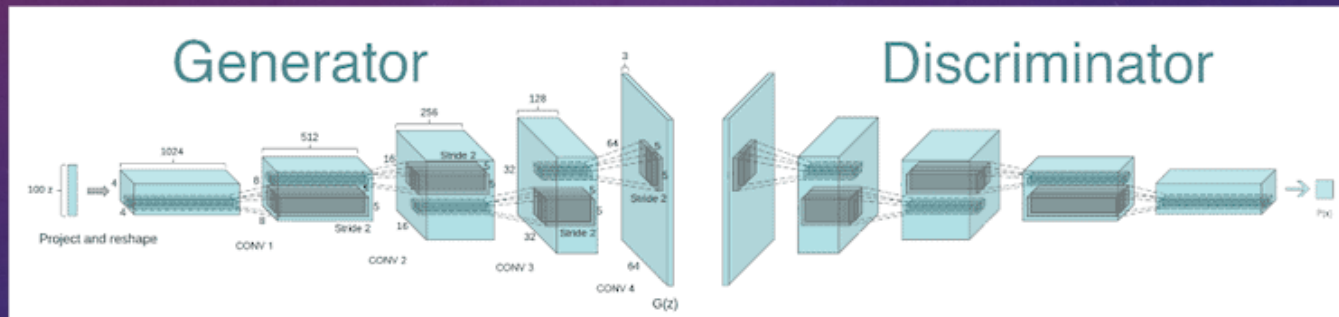
APPRENTISSAGE PROFOND EN TAP



- Réseaux de neurones profonds (Deep Neural Networks)
 - Modèles complexes
 - Grande quantité d'information pour l'apprentissage
 - Organisation interne proche de l'organisation du cerveau humain (Nagamine, 2015; Chang, 2016)

FUTUR RAP ?

- Architectures de bout en bout (End2End)
 - Connectionist Temporal Classification (CTC)
 - Modèle de l'attention
 - Transducteurs
 - Transformeur
- Réseaux Antagonistes Génératifs (GAN)



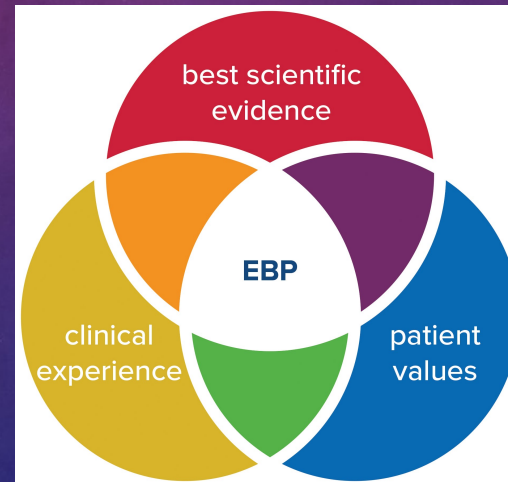
TAP EN CLINIQUE

(Balaguer, 2021)

- Évaluation en clinique courante : aide aux choix des stratégies thérapeutiques
- Evidence Based Practice (Sackett et al., 1996)

- Outils d'évaluation
 - Peu sont disponibles pour la parole, surtout en oncologie
 - Mesure très largement perceptive (Middag et al., 2014; Pommée et al., 2021)
- Intérêt d'analyses acoustiques et automatiques pour pallier les biais perceptifs (Van Nuffelen et al., 2009; Fex, 1992)
- Champ de recherche en plein essor : C2SI, RUGBI, DAPADAF-E, consortium TAPAS...

- État des lieux (Pommée et al., 2021)
 - ¼ des experts cliniciens n'évaluent pas l'impact fonctionnel du trouble de parole
 - Inadéquation des questionnaires utilisés (VHI = 70 %)

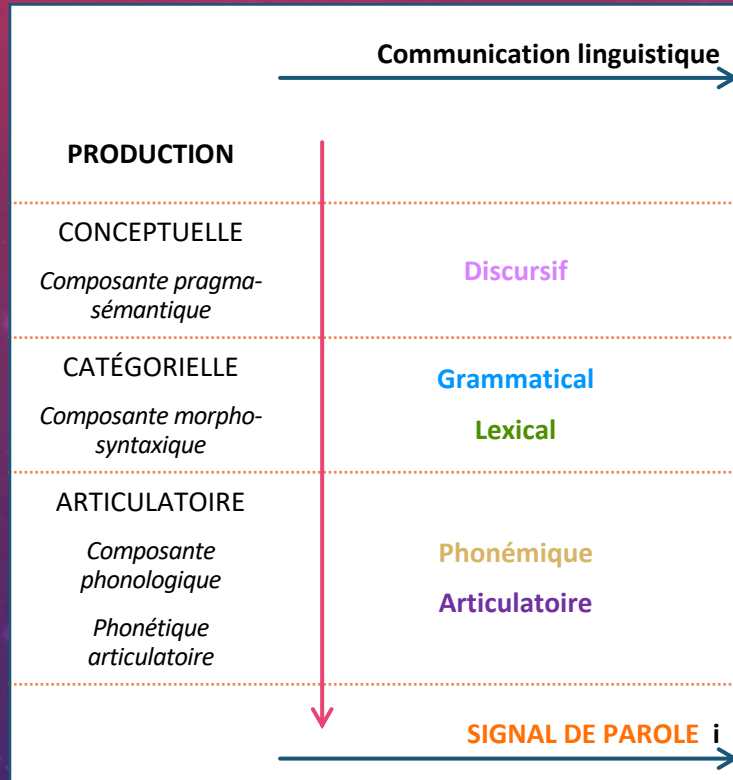


- Étude préliminaire (11 patients) :
 - 6/11 auraient aimé que les conséquences du trouble de parole soient davantage prises en compte
 - 2/11 considèrent qu'elles n'ont pas assez été prises en compte

EXEMPLE TAP

(Balaguer, 2021)

(Caron, 1989)



« non j'appelle pas souvent »

Signal de parole (Praat) :

- Macrovariabilité de l'intensité et de F0
- Spectre : pente et inclinaison de la courbe
- Cepstre : prééminence du pic cepstral lissé

Articulatoire – Phonétique articulatoire :

- Mesures temporelles de durée de parole (*WebRTC-VAD, Praat*)
- Débit de parole et d'articulation (*Praat*)
- Segments vocaliques (*Speechtools*)

Articulatoire – Composante phonologique (TDNNF-HMM) :

- Inventaire phonémique
- Score de confiance

Composante morpho-syntaxique – Lexical (TDNNF-HMM) :

- Inventaire lexical
- Scores de confiance

1	CFC12-ECV-16k_mono 1	848.54	0.23	le	0.52
2	CFC12-ECV-16k_mono 1	848.97	0.30	club	0.27
3	CFC12-ECV-16k_mono 1	849.39	0.24	pas	0.79
4	CFC12-ECV-16k_mono 1	849.63	0.27	vont	0.43

- Caractéristiques lexicales (*lexique.org*)
- Diversité et densité lexicales

Articulatoire – Composante phonologique (TDNNF-HMM) :

- 1 CFC12-ECV-16k_mono_848_2499999994703_850_559999
- 9994682 L E~ L P A E P a S B

Composante morpho-syntaxique – Grammatical :

- Classes grammaticales (*lexique.org*)
- Formes actives et supplémentaires (*Iramuteq*)

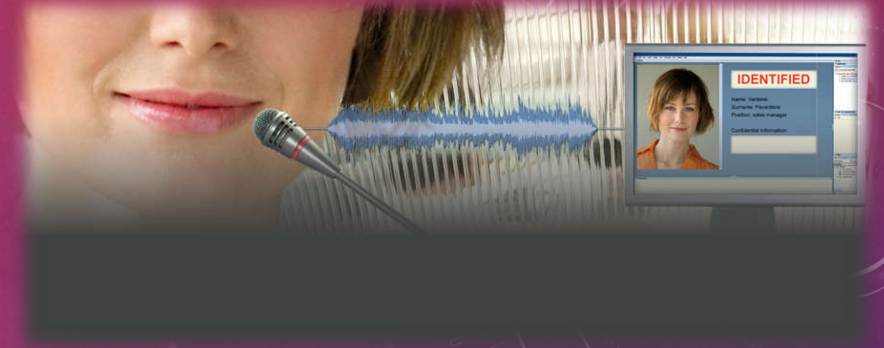
Composante pragma-sémantique – Discursif :

- Analyse thématique (*Iramuteq*)
- Analyse de sentiment (*Microsoft Azure*)



DE LA RECONNAISSANCE DE LA PAROLE À LA RECONNAISSANCE DU LOCUTEUR

LA RECONNAISSANCE DU LOCUTEUR



- Recherche à identifier les caractéristiques particulières d'un locuteur
- Activité orthogonale à la reconnaissance de la parole
- Principe
 - ✓ Extraction de paramètres acoustiques à partir de la voix du locuteur
 - ✓ Apprentissage et modélisation de ces caractéristiques
 - ✓ Identification, vérification :
on confronte les caractéristiques d'une nouvelle voix aux modèles des locuteurs appris

L'IDENTIFICATION VOCALE DANS UN CADRE CRIMINALISTIQUE

- Utilisation de techniques de reconnaissance du locuteur dans le cadre d'affaires criminelles
- La police scientifique possède un enregistrement du criminel (inconnu) + des enregistrements d'un ou plusieurs suspects : l'identification de la personne est-elle possible ?
 - ✓ Par des humains ?
 - ✓ Par des machines ?
- Principe de précaution :
 - ✓ la voix n'est ni une empreinte digitale (empreinte vocale), ni stable comme l'ADN
 - ✓ La voix est instable à court terme, à long terme, sujette au contexte...
 - ✓ La voix est modifiable



COMPARAISON DE VOIX APPLIQUÉE À LA CRIMINALISTIQUE : LE PROJET voxcrim

- Projet financé par l'Agence Nationale de la Recherche
- Interdisciplinaire
 - ✓ Laboratoire d'informatique (LIA, Avignon)
 - ✓ Laboratoire de phonétique (LPL, Aix ; LPP, Paris)
 - ✓ Laboratoire de police technique et scientifique ; Institut de recherche criminelle de la Gendarmerie nationale
 - ✓ Laboratoire National d'Essai
- Objectifs :
 - ✓ Étudier les performances et limites de la perception humaine
 - ✓ Étudier les performances et limites de l'identification par ordinateur
 - ✓ Proposer un cadre méthodologique
 - ✓ Mettre en place une accréditation liée à une formation



UNE ANALOGIE A FAIRE AVEC LA LARYNGOPHONIATRIE ?

- Les limites de la perception humaine ?

- ✓ La perception de la voix est fiable dans certaines conditions mais elle peut nous induire en erreur
- ✓ « on entend ce qu'on s'attend à entendre, à ce qu'on veut entendre »
- ✓ Nécessité de blind test

> J Voice. 2013 Jul;27(4):481-5. doi: 10.1016/j.jvoice.2013.03.015. 

Top-down mechanisms in dysphonia perception: the need for blind tests

Alain Ghio¹, Joana Révis, Sabine Merienne, Antoine Giovanni

Affiliations + expand

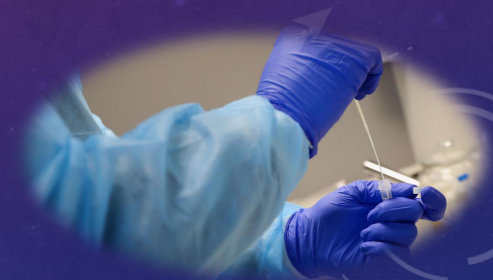
PMID: 23809570  DOI: 10.1016/j.jvoice.2013.03.015 

- Un cadre méthodologique

- ✓ La notion de prélèvement de voix
 - Aspects techniques (microphone, enregistreur...), linguistiques (quel corpus ? Quelles consignes ?)
 - Si le prélèvement n'est pas bien fait, il ne peut être reçu comme un élément de preuve

- Formation et accréditation

- ✓ Seul un personnel formé et accrédité est autorisé à faire un prélèvement et une analyse
- ✓ Seules les méthodologies testées et éprouvées sont autorisées à être utilisées



LA RECONNAISSANCE DU LOCUTEUR EN LARYNGOPHONIATRIE

- On modélise non pas un locuteur mais un groupe de locuteurs
 - ✓ Ex : G0= voix normale, G1 = dysphonies légères, G2 = dysphonies moyennes, G3= dysphonies sévères
- La voix d'un nouveau patient est comparée aux modèles via un système automatique
- Le système automatique désigne un groupe avec des probabilités d'appartenance
 - *M. Wester, Automatic classification of voice quality: Comparing regression models and hidden markov models, in: VOICEDATA98, Symposium on Databases in Voice Quality Research and Education, 1998, pp. 92-97*
 - *Fredouille, Pouchoulin, Bonastre, Azzarello, Giovanni, Ghio. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). Interspeech, 2005, Lisboa, France. pp.149-152.*
- Besoin d'importantes bases de données
- Données bien catégorisées

Test Gr.	Classification system Response			
	0	1	2	3
0	19	1	0	0
1	3	14	2	1
2	2	7	9	2
3	0	0	7	13

DU DECODAGE À L'ENCODAGE

QUID DE LA SYNTHÈSE VOCALE ?



LA SYNTHÈSE VOCALE ET LA LARYNGOPHONIATRIE

- Pour un usage en voix de substitution
 - ✓ Patients laryngectomisés, paralysés...
- La synthèse text-to-speech est une question résolue de façon générale
 - ✓ Exemple : www.acapela-group.com, voxygen.fr
 - ✓ Fondée sur de l'apprentissage par l'exemple
- Possibilité de personnalisation
- Quid sur des patients atteints de troubles de la parole ?
 - ✓ il faut enregistrer la voix du patient en bonne santé pour l'utiliser en synthèse
 - ✓ Précaution rarement appliquée

LA PROSODIE, TALON D'ACHILLE DE LA SYNTHÈSE

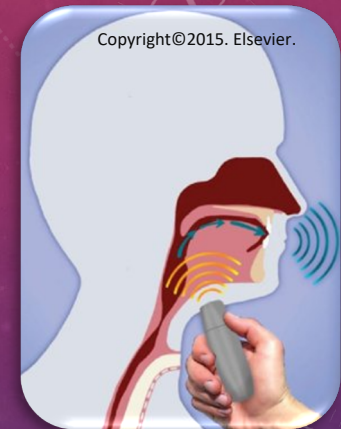
- La prosodie = mélodie et rythme de la parole
- Actuellement, pas toujours bien synthétisée
 - ✓ Accentue l'effet artificiel
 - ✓ Une limite à la transmission des émotions, de la personnalité...
- Un vrai problème dans les langues à tons comme le chinois
 - ✓ Perte d'intelligibilité si les tons ne sont pas produits



<http://www.phonetics.ucla.edu/vowels/chapter2/cantonese>

CANTONESE si			
Chinese Character	Tone symbol	Tone description	English gloss
詩	˥ 55	highlevel	'poem'
試	˨˨ 33	mid level	'to try'
事	˨˨ 22	low level	'matter'
時	˨˥ 21	low falling	'time'
使	˨˥ 24	low high	'to cause'
市	˨˥ 23	low mid	'city'

LES VOIX PROTHÉTIQUES, ELECTROLARYNX



- Electrolarynx classique
 - ✓ Monotone
 - ✓ Et donc robotique
- Electrolarynx à fréquence vibratoire ajustable
- Permet de produire des variations de fréquence fondamentale

Case Reports > Med Biol Eng Comput. 2017 Aug;55(8):1463-1472.

doi: 10.1007/s11517-016-1606-6.  Epub 2016 Dec 24.

Development and evaluation of wheel-controlled pitch-adjustable electrolarynx

Li Wang¹, Yijun Feng¹, Ze Yang¹, Haijun Niu²

Affiliations + expand

PMID: 28013472  DOI: 10.1007/s11517-016-1606-6 

NOS PROJETS TECHNOLOGIQUES

• ANR RUGBI (2019-2023)

Design and Development of a Speech Intelligibility Test Based on Pseudowords in French: Why and How?

Muriel Lalain, Alain Ghio, Laurence Giusti, Danièle Robert, Corinne Fredouille and Virginie Woisard

https://doi.org/10.1044/2020_JSLHR-19-00088

Lang Resources & Evaluation
<https://doi.org/10.1007/s10579-020-09496-3>

ORIGINAL PAPER

C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers

Virginie Woisard¹ • Corine Astésano² • Mathieu Balaguer¹ • Jérôme Farinas³ • Corinne Fredouille¹ • Pascal Gaillard² • Alain Ghio³ • Laurence Giusti³ • Imed Laaridh⁵ • Muriel Lalain³ • Benoît Lepage² • Julie Maclair⁵ • Olivier Nocaudie² • Julien Pinquier² • Gilles Pouchoulin³ • Michèle Puech¹ • Danièle Robert² • Vincent Roger²

© Springer Nature B.V. 2020

Du recueil à l'exploitation des corpus de parole "pathologique": comment accéder à la variation physiopathologique?

Alain Ghio, Gilles Pouchoulin, François Viallet, Antoine Giovanni, Virginie Woisard, Lise Crevier-Buchman, Fabrice Hirsch, Camille Fauth, Corinne Fredouille

► To cite this version:

Alain Ghio, Gilles Pouchoulin, François Viallet, Antoine Giovanni, Virginie Woisard, et al.. Du recueil à l'exploitation des corpus de parole "pathologique": comment accéder à la variation physiopathologique?. Corpus, Bases, Corpus, Langage - UMR 7320, 2021, 10.4000/corpus.3677. hal-03145102

Construction d'un index holistique d'impact sur la communication des troubles de la parole chez des patients traités pour un cancer oral ou oropharyngé
Mathieu Balaguer¹, Julien Pinquier¹, Jérôme Farinas¹, Benoît Lepage¹, Virginie Woisard¹

Introduction

Puis d'être généraliser l'évaluation de l'impact de trouble de parole en oncologie ORL (L2)

- Questionnaires évaluant les limitations
- Qualité et des restrictions de participation: PIR (2)
- D'autres évaluer la dynamique communicationnelle, non en tant que variable en oncologie ORL (KOWI) DP (R)

Objectif

Construire un score holistique permettant de quantifier le niveau l'impact fonctionnel sur la communication des troubles de parole ORL de patients traités pour un cancer de la cavité buccale ou de l'oropharynx.

Matériel et méthode

35 patients (95 % hommes, âge médian = 67 ans, EO 12) traités pour un cancer oral (48 %) ou oropharyngé (40 %) (six localisations: 4 %)

Comparés par les patients de 3 auto-questionnaires

- Questionnaire communicationnel (C2SI) (R) (2)
- Impact psychosocial et trouble de parole (IMP) (R)
- Niveau de parole (NVP) (R) (2)
- Qualité de vie globale et relative à la parole (EORTC QLQ-C15 et QLQ-C15P)

Méthodologie de réduction de dimensionnalité

- Validité d'explorer: choix des items pertinents des items des questionnaires par un critère d'apriori
- Validité de construit: calcul des items redondants et l'impact sur variables (analyse de corrélation inter-items)
- Validité de contenu: réduction de la charge latente

Résultats

- 203 items issus des questionnaires sont retenus à la communication au sens large (oral, écrit, vécu personnel...)
- Validité d'explorer: appui sur un comité de 9 experts (six des disciplines sciences du langage, médecine, informatique et traitement du signal, orthophonie)
- Définition de la communication dans le contexte de cette étude: relative à l'expression orale, notion d'interaction relative traduite avec un item
- Sur les 25 patients: n = 8,79 items ou score holistique et le score préexistant de sévérité de trouble de parole
- Sur les 25 patients: n = 8,79 items ou score holistique et le score préexistant de sévérité de trouble de parole
- Validité de construit: sélection statistique des items redondants avec exclusion de l'item le moins varié dans les bases considérées (n = 20)
- 40 items restants (91 % retenus)
- Construction du score holistique:
 - Analyse factorielle PCF
 - Charges des facteurs supérieures à 0,40
 - Impact psychosocial et trouble de parole (IMP) (R)
 - Niveau de parole (NVP) (R)
 - Valeur seuil de l'item 1 = 17,00

Discussion

Score holistique permettant de quantifier l'impact de ce type d'états en oncologie ORL, à compléter par l'item 1

- La facilité temporelle par la composition des items retenus dans la construction du score à deux moments proches par les mêmes patients
- La validation externe sur un nouvel échantillon de patients

Nouvelles perspectives quant à l'étude de la qualité de vie après traitement d'un cancer oral ou oropharyngé, dont l'évaluation des capacités communicationnelles, peut être une des composantes.

Références

1. Balaguer M, Pinquier J, Farinas J, Lepage B, Woisard V, et al. (2021) Construction d'un index holistique d'impact sur la communication des troubles de la parole chez des patients traités pour un cancer oral ou oropharyngé. *Corpus, Bases, Corpus, Langage - UMR 7320*, 2021, 10.4000/corpus.3677. hal-03145102

Une méthode automatique non supervisée pour évaluer le score de sévérité de la parole chez les patients traités pour un cancer ORL

Vincent Roger, Jérôme Farinas, Virginie Woisard, Julien Pinquier

► To cite this version:

Vincent Roger, Jérôme Farinas, Virginie Woisard, Julien Pinquier. Une méthode automatique non supervisée pour évaluer le score de sévérité de la parole chez les patients traités pour un cancer ORL. Séminaire AFCP 2021 - Phonétique Clinique, May 2021, Toulouse (virtuel), France. 2021. hal-03269228

INTERSPEECH 2020
October 25–29, 2020, Shanghai, China



Towards Interpreting Deep Learning Models to Understand Loss of Speech Intelligibility in Speech Disorders

Step 1: CNN model-based phone classification

Sondes Abderrazek¹, Corinne Fredouille¹, Alain Ghio², Muriel Lalain³, Christine Meunier², Virginie Woisard²

¹LIA, Avignon University, France

²Aix-Marseille Univ, LPL, CNRS, Aix-en-Provence, France

³UTJ, Octogone-Lordat, Toulouse University & Toulouse Hospital, France

(sondes.abderrazek, corinne.fredouille)@univ-avignon.fr

Automatic extraction of speech rhythm descriptors for speech intelligibility assessment in the context of Head and Neck Cancers

Robin Vaysse, Jérôme Farinas, Corine Astésano, Régine André-Obrecht

► To cite this version:

Robin Vaysse, Jérôme Farinas, Corine Astésano, Régine André-Obrecht. Automatic extraction of speech rhythm descriptors for speech intelligibility assessment in the context of Head and Neck Cancers. INTERSPEECH 2021, ISCA: International Speech and Communication Association, Aug 2021, Brno, Czech Republic. hal-03269227

• ANR PHLES-NID (2022-2026)

Dysphagia 23:102-109 (2008)
DOI: 10.1007/s00455-007-9099-y

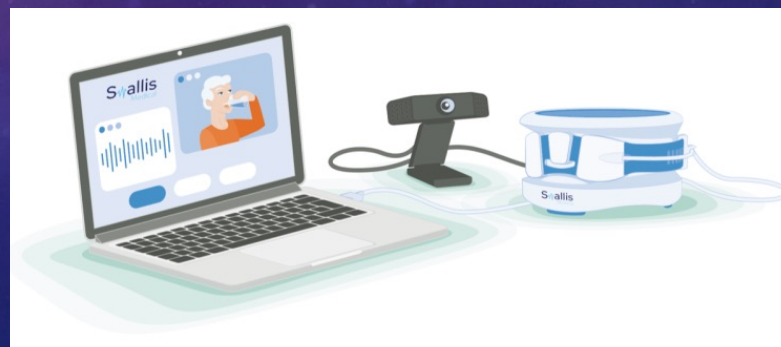
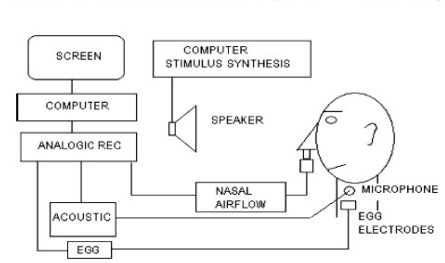
Dysphagia
© Springer Science+Business Media, LLC 2008

Use of Reaction Time in the Temporal Analysis of Normal Swallowing

Bernard Roubeau, PhD,¹ Sylvain Morinière, MD, PhD,² Sophie Périé, MD, PhD,¹ Anne Martineau, SLT,¹ Jannic Falières, SLT,¹ and Jean Lacau St Guily, MD¹

¹Department of Otolaryngology-Head and Neck Surgery, Faculty of Medicine University Pierre & Marie Curie, Tenon Hospital, Paris 6, Paris, France and ²Department of Otolaryngology-Head and Neck Surgery, Faculty of Medicine University François Rabelais, Bretonneau Hospital, Tours, France

B. Roubeau et al.: Reaction Time in Analysis of Normal Swallowing



NOS PROJETS TECHNOLOGIQUES

- Plateforme de Parole Atypique (<https://paty.irit.fr/demo/>)



- PHRIP DAPADAF-E (2020-2024) : *Validité d'une tâche de Décodage Acoustico Phonétique Automatisée sur les Déficiets Anatomo-Fonctionnels dans l'Évaluation paramédicale des troubles de la parole des patients traités pour un cancer de la cavité buccale ou de l'oropharynx*
- Réseau européen TAPAS (2017-2021) : Training Network on Automatic Processing of PAtHological Speech

MERCI!

JEROME.FARINAS@IRIT.FR

ALAIN.GHIO@LPL-AIX.FR



CORINNE FREDOUILLE
(LIA, UNIV. AVIGNON)



GILLES POUCHOULIN
(LPL, UNIV. AIX MARSEILLE)



VIRGINIE WOISARD
(CHU TOULOUSE, UNIV. TOULOUSE)




MATHIEU BALAGUER
(IRIT, UNIV. TOULOUSE & CHU)



ANTOINE GIOVANNI
(CHU MARSEILLE, AIX-MARSEILLE UNIV)

IRIT
Institut de Recherche
en Informatique de Toulouse



Mardi 12 Octobre 2021
14h00
UT3 Paul Sabatier, IRIT, Auditorium J. Herbrand

Mathieu BALAGUER
Equipe SAMOVA, IRIT

Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé

Jury :
- Emmanuel Bégin, Professeur des Universités - Praticien Hospitalier, CHU Caen, Unité INSERM U1086 ANTICYPE - Cancéris et prévention - rapporteur
- Nathalie Bernard-Schostalski, Directrice de Recherche, Laboratoire GIPSA Lab (Unité mixte du CNRS, de Grenoble-INP et de l'Université de Grenoble-Alpes) - rapporteur
- Ruchika Singh, Professeur des Universités, LIPSA, Université de Strasbourg - examinateur
- Jérôme Farinas, Maître de Conférences, Université Toulouse III Paul Sabatier - co-encadrant de thèse
- Julien Pradier, Maître de Conférences, Université Toulouse III Paul Sabatier - Directeur de thèse
- Virginie Woisard, Praticienne Hospitalière, CHU Toulouse, Laboratoire LNPL, Université Jean-Jaures Toulouse 8 - Co-encadrant de thèse

Résumé : Les troubles de parole sont une problématique fréquemment rencontrée après traitement d'un cancer de la cavité buccale ou de l'oropharynx, mais peu d'études s'intéressent à l'heure actuelle aux conséquences de ce trouble sur les capacités de communication des patients ou leur qualité de vie. Or, en contexte clinique, l'optimisation des capacités de communication est un objectif thérapeutique majeur dans le suivi de ces patients. En pratique courante, l'évaluation des troubles de parole est réalisée majoritairement de façon perceptive et donne des scores prédisant mal l'impact de ces troubles sur la communication. Le développement des outils d'analyse automatique de la parole permet de pallier la variabilité des scores liée au caractère perceptif de l'évaluation classique.

Dans cette thèse, nous avons cherché à mesurer l'altération de la communication au moyen d'analyses automatiques de la parole spontanée.

Nous avons étudié trois aspects : la mesure de l'altération de la communication, l'analyse automatique de la parole spontanée, et la prédiction de l'altération de la communication par les paramètres automatiques. Pour ce faire, nous avons constitué un nouveau corpus de parole auprès de 25 sujets traités pour un cancer de la cavité buccale ou de l'oropharynx. Il comprend une tâche de parole spontanée, enregistrée au cours d'un entretien semi-dirigé, mais aussi des questionnaires autorisant la mesure de la communication et des facteurs associés à la parole et à la communication.

Concernant le premier aspect, un score de référence, mesurant de façon holistique la communication, a été construit, permettant de combler le manque d'outils disponibles en oncologie ORL pour cette mesure.

Pour le deuxième aspect, concernant l'analyse automatique de la parole, une revue systématique de littérature nous a conduits à nous intéresser aux outils applicables à l'analyse de la parole spontanée, qui est le contexte de production de parole le plus proche de la communication quotidienne. Puis un processus de sélection a abouti à retenir 75 paramètres automatiques pertinents et non redondants.

Soutenance de Thèse

05 61 55 65 10
info@irit.fr
www.irit.fr

