



HAL
open science

La laryngophoniatrie du futur

Alain Ghio, Jérôme Farinas

► **To cite this version:**

Alain Ghio, Jérôme Farinas. La laryngophoniatrie du futur. 127ème Congrès National de la Société Française d'ORL (SFORL 2021), 2021, Paris, France. hal-03365266v2

HAL Id: hal-03365266

<https://hal.science/hal-03365266v2>

Submitted on 8 Oct 2021 (v2), last revised 9 Oct 2021 (v3)

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.

LA LARYNGOPHONIATRIE DU FUTUR



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

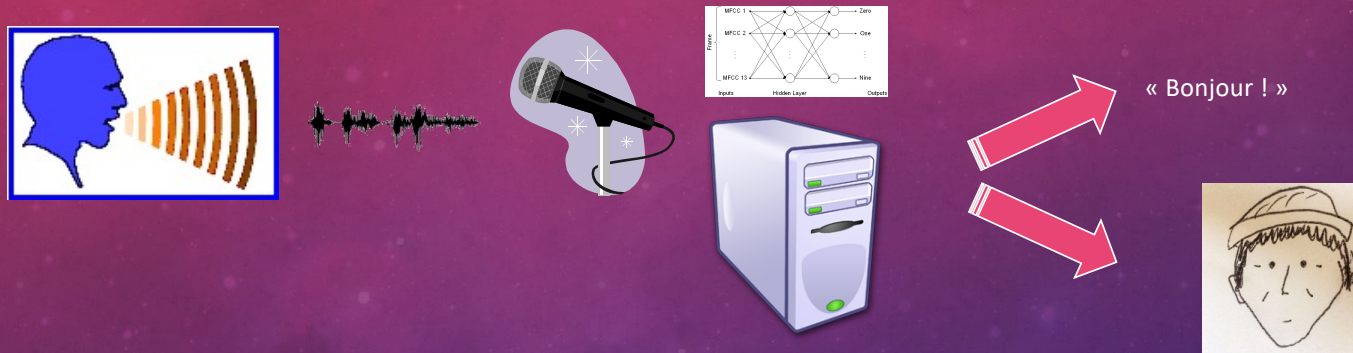


ALAIN GHIO, LABORATOIRE PAROLE ET LANGAGE, CNRS, AIX-MARSEILLE UNIVERSITE

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

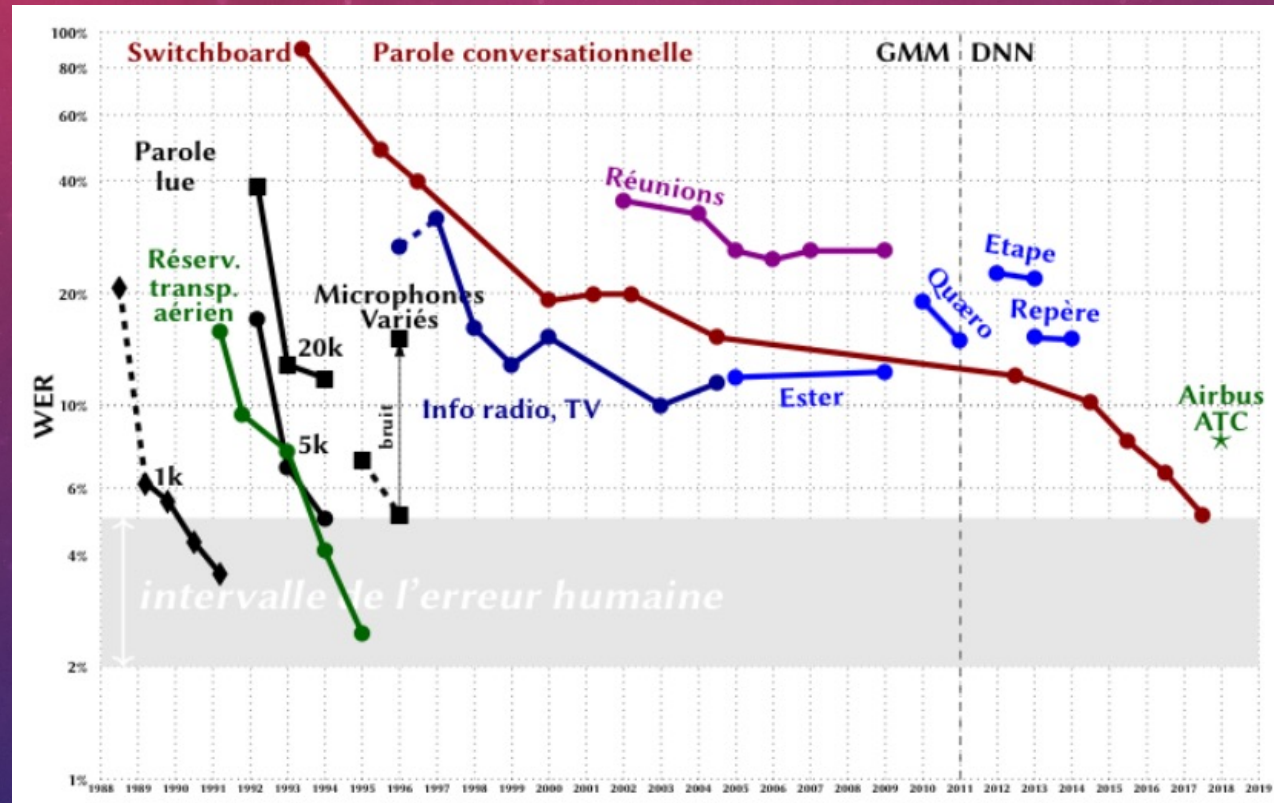
POURQUOI LE TAP?

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

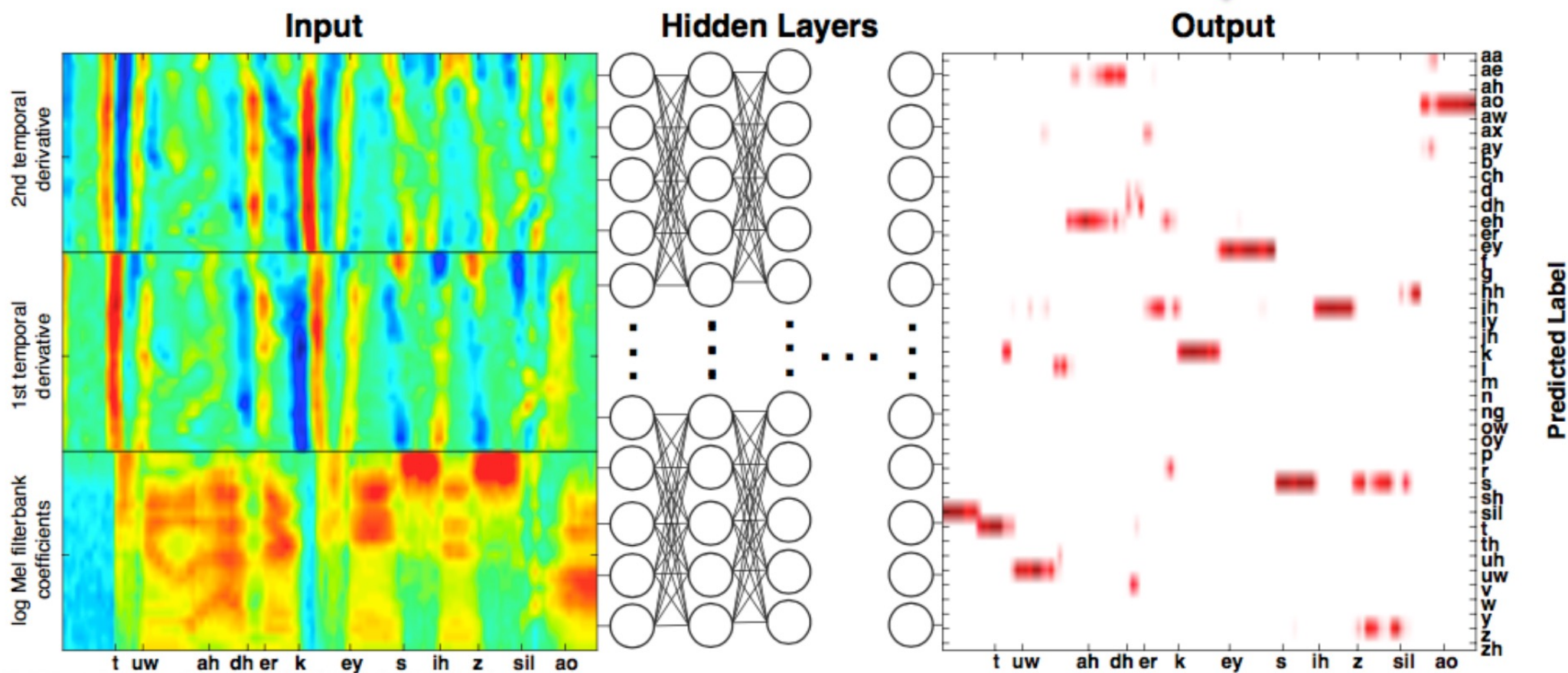
RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

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

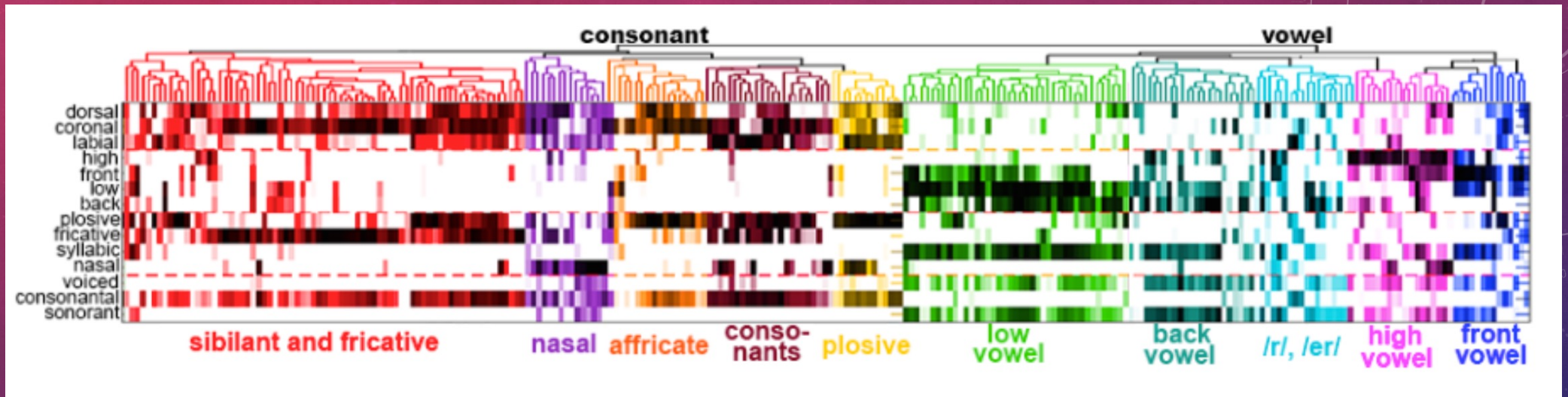
QUOI DE NEUF EN RAP ?



APPRENTISSAGE PROFOND EN RAP



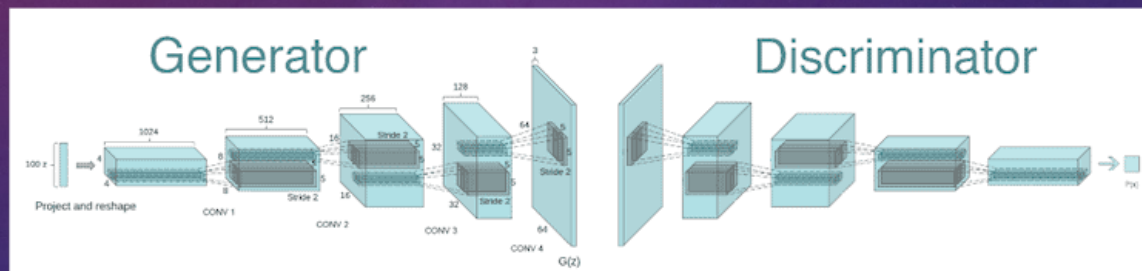
APPRENTISSAGE PROFOND EN TAP



- 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)



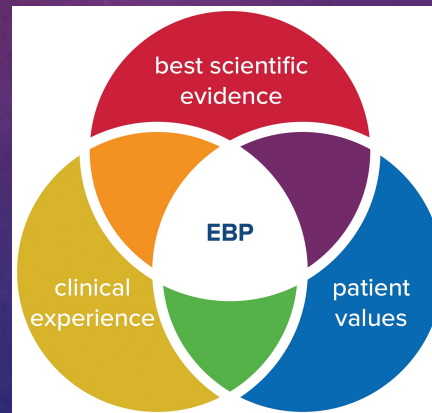
TAP EN CLINIQUE

(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 cancérologie
 - 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 %)

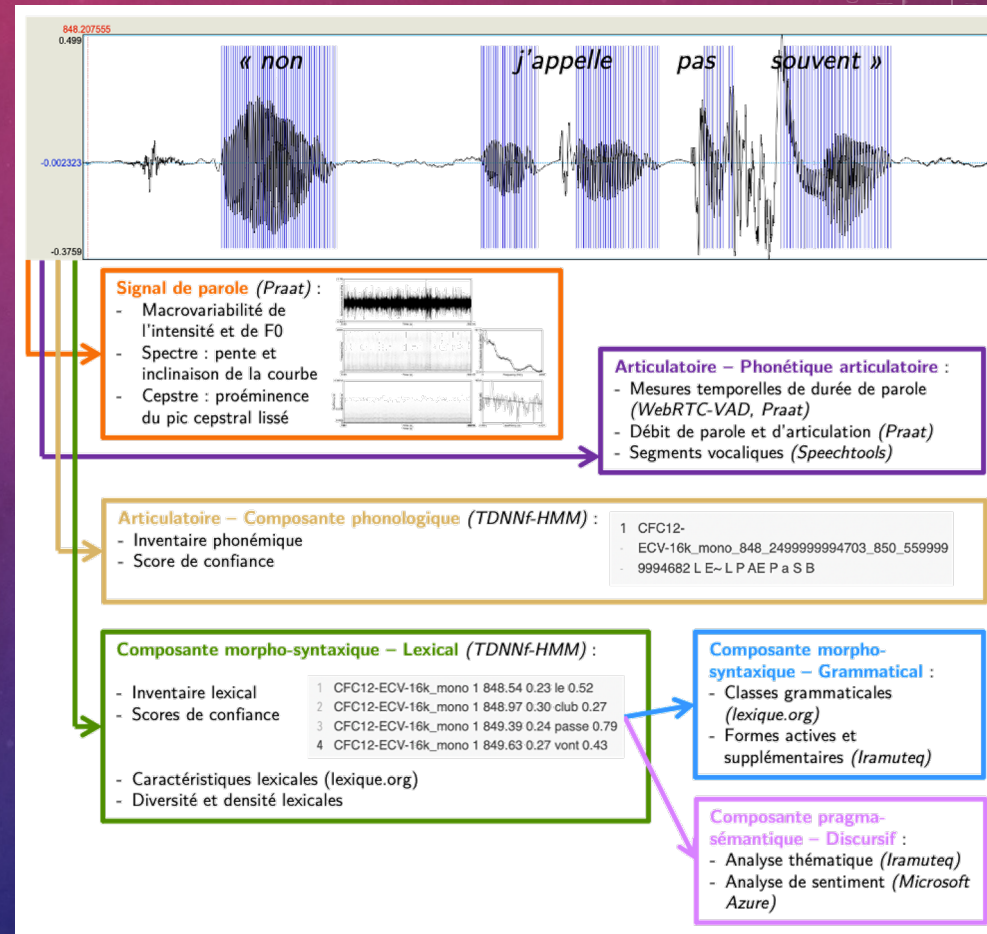
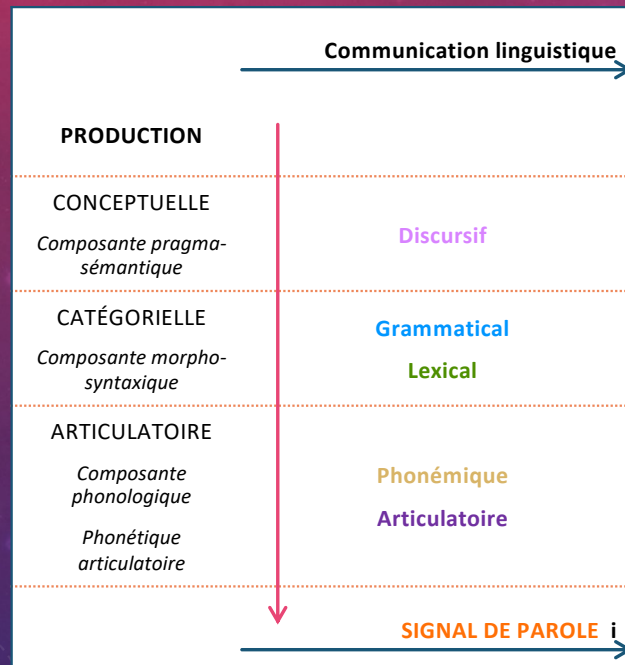


- Étude préliminaire (11 patients) :
 - 6/11 auraient aimé que les conséquences du trouble de parole soient davantage prises en compte
 - 2/11 considèrent qu'elles n'ont pas assez été prises en compte

EXEMPLE TAP

(Balaguer, 2021)

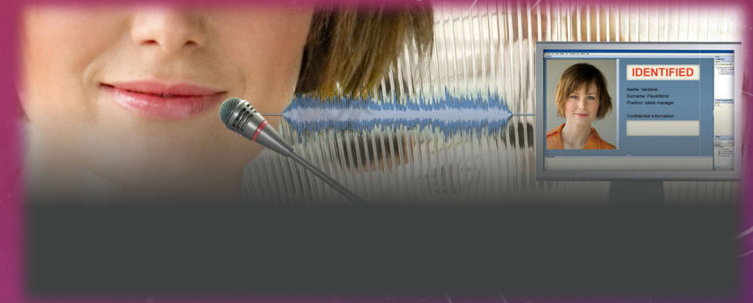
(Caron, 1989)





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

- 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

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

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

Alain Ghio¹, Joana Révis, Sabine Merienne, Antoine Giovanni

Affiliations [+ expand](#)

PMID: 23809570 [CR](#) DOI: 10.1016/j.jvoice.2013.03.015 [CR](#)



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 VoiceQuality 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

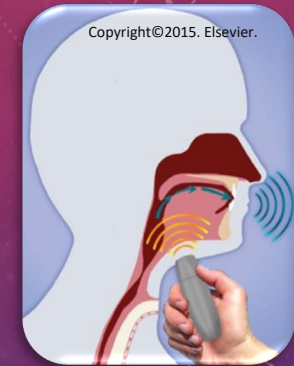
- La prosodie = mélodie et rythme de la parole
- Actuellement, pas toujours bien synthétisée
 - ✓ Accentue l'effet artificiel
 - ✓ Une limite à la transmission des émotions, de la personnalité...
- Un vrai problème dans les langues à tons comme le chinois
 - ✓ Perte d'intelligibilité si les tons ne sont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



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

CANTONESE si			
Chinese Character	Tone symbol	Tone description	English gloss
詩	7 55	high level	'poem'
試	4 33	mid level	'to try'
事	4 22	low level	'matter'
時	4 21	low falling	'time'
使	4 24	low high	'to cause'
市	4 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 Wang ¹, Yijun Feng ¹, Ze Yang ¹, Haijun Niu ²

Affiliations + expand

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

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 Déficiences Anatomo-Fonctionnelles 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)



VIRGINIE WOISARD
(CHU TOULOUSE, UNIV. TOULOUSE)



GILLES POUCHOULIN
(LPL, UNIV. AIX MARSEILLE)




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 Babin, Professeur des Universités - Praticien Hospitalier, CHU Caen, UNIV. NEMERUS UTORB ANTOCIFE - Cancer
- Emmanuel Babin, Professeur des Universités - Praticien Hospitalier, Laboratoire GPTSA Lab. UNIV. NEMERUS ANTOCIFE - Cancer
- Université de Grenoble Alpes - Approche
- Raphaël Foni, Professeur des Universités, Université de Bourgogne, Bourgogne
- Jérôme Farinas, Maître de Conférences, Université Toulouse III Paul Sabatier - Co-encadrant de thèse
- Alain Ghio, Maître de Conférences, Université Toulouse III Paul Sabatier - Directeur de thèse
- Virginie Woisard, Praticienne Hospitalière, CHU Toulouse, Laboratoire LPL, Université Jean Jaurès Toulouse II - Co-encadrante 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 préjudicant 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 après de 28 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.

05 61 55 85 10
info@irit.fr
www.irit.fr

