HAL
open science

Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique

Alain Ghio, Muriel Lalain, Laurence Giusti, Gilles Pouchoulin, Danièle Robert, Marie Rebourg, Corinne Fredouille, Imed Laaridh, Virginie Woisard

To cite this version:

Alain Ghio, Muriel Lalain, Laurence Giusti, Gilles Pouchoulin, Danièle Robert, et al.. Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique. XXXIIe Journées d'Etudes sur la Parole, LPL, 2018, Aix-en-Provence, France. pp.285-293, 10.21437/jep.2018-33 . hal-01770161v2

HAL Id: hal-01770161

https://hal.science/hal-01770161v2

Submitted on 9 Jan 2019

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.

Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique

Alain Ghio[1], Muriel Lalain[1], Laurence Giusti[1], Gilles Pouchoulin[1], Danièle Robert[1,2], Marie Rebourg[1], Corinne Fredouille[3], Imed Laaridh[3], Virginie Woisard[4]
(1) Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France
(2) Service ORL, APHM, Marseille, France
(3) Laboratoire d'Informatique d'Avignon, Avignon, France
(4) Service ORL, CHU Larrey, URI Octogone-Lordat, Toulouse, France

alain.ghio@lpl-aix.fr

RESUME

Les limitations actuelles des tests d'intelligibilité effectués sur des locuteurs ayant une production atypique de la parole résident dans la capacité des auditeurs à restaurer les séquences distordues. Le résultat est une mesure surévaluée par rapport à la performance articulatoire réelle. Nous présentons un test d'intelligibilité fondé sur la prononciation de pseudo-mots de façon à complètement neutraliser les effets de lexicalité ou d'apprentissage des items par les auditeurs.
126 locuteurs (41 sujets sains et 85 patients atteints de troubles de la parole) ont produit chacun 52 pseudo-mots tirés aléatoirement d'une liste de 89346 formes possibles. 40 auditeurs ont retranscrit ces productions. Les transcriptions orthographiques ont été phonétisées puis comparées aux formes phonétiques attendues par un algorithme de Wagner-Fischer qui intègre les phénomènes d'insertion, élision et substitution de phonèmes. Les résultats montrent que les formes perçues chez les patients sont en moyenne à une distance bien plus élevée que chez les sujets contrôles.

MOTS-CLES : Intelligibilité; parole atypique; traits phonétiques ; décodage acoustico-phonétique

ABSTRACT

A measure of intelligibility by acoustic-phonetic decoding of pseudo-words in the case of atypical speech

The current intelligibility tests performed on speakers with atypical speech production are limited by the ability of listeners to restore distorted sequences. The result is an overvalued measure compared to the actual articulatory performance. We present an intelligibility test based on the pronunciation of pseudo-words in order to neutralize unwanted lexical and learning effects of items by the listeners. 126 speakers (41 healthy subjects and 85 patients) each produced 52 pseudo-words randomly drawn from a list of 89346 possible forms. 40 listeners have transcribed these productions. Orthographic transcriptions were phonetized and compared to the phonetic forms expected by a Wagner-Fischer algorithm that integrates the phenomena of insertion, elision and phoneme substitution. The results show that the forms perceived with patients are on average at a greater distance than with healthy subjects.

KEYWORDS: Intelligibility; atypical speech; phonetic features ; acoustic-phonetic decoding

10.21437/JEP.2018-33

1 Pourquoi une mesure d'intelligibilité sous forme de décodage acoustico-phonétique ?

1.1 Intelligibilité et compréhensibilité de la parole

La perception de la parole est un processus complexe qui intègre à la fois un flux ascendant d'informations provenant du signal vocal mais aussi un flux descendant fondé sur les informations de haut niveau détenues par l'auditeur. Le flux ascendant (« bottom-up ») est principalement une opération de décodage acoustico-phonétique qui consiste à identifier les phonèmes à partir du signal de parole. Les phonèmes, pouvant être considérés comme les plus petites unités permettant d'opposer du sens, sont les éléments de base de l'intelligibilité du discours, c'est-à-dire du degré de précision avec lequel le message est compris par l'auditeur. Le décodage acoustico-phonétique est donc le processus fondamental pour mesurer perceptivement l'intelligibilité d'un locuteur.

Le flux descendant (« top-down ») fait appel chez l'auditeur à un ensemble d'informations qu'il détient à différents niveaux : la connaissance du lexique de façon générale, la connaissance du contexte de la situation de communication pouvant potentiellement restreindre considérablement le lexique de circonstance, la connaissance des communicants... De ce fait, lorsqu'un auditeur entend un énoncé dégradé, bruité ou phonétiquement appauvri, ces processus top-down entrent en jeu pour restaurer ce qui est distordu et optimiser l'intelligibilité du message (Warren et al., 1970). Les effets de lexicalité, c'est-à-dire le fait qu'une séquence sonore ou orthographique fasse référence à un mot de notre vocabulaire, sont notamment très forts. Les travaux de (Ganong, 1980) ont montré qu'en anglais, un son phonétiquement ambigu t/d sera préférentiellement perçu [d] s'il est placé devant une séquence [aʃ] en référence au mot « dash », et inversement, le même son sera perçu [t] devant une séquence [ask] en référence au mot « task ». Il faut remarquer qu'en français, le résultat serait inversé : un son phonétiquement ambigu t/d sera préférentiellement perçu [t] devant une séquence [aʃ] en référence au mot « tache » mais il sera perçu [d] s'il est placé devant une séquence [isk] en référence au mot « disque ». A ces effets de lexicalité s'ajoutent d'autres phénomènes comme la fréquence des mots (les mots usuels sont plus facilement reconnus), les règles phonotactiques de la langue (une séquence [vrsitʃ] est peu probable en français), le savoir partagé relatif au contexte de la conversation.

Dans le cas où nous nous intéressons à l'intelligibilité d'un locuteur produisant une parole atypique (production pathologique de la parole, apprentissage des langues, acquisition ou vieillissement), ces mécanismes top-down peuvent s'avérer gênants pour mesurer le degré de précision/perturbation dans la mesure où ils interviennent chez l'auditeur de façon variable et qu'ils peuvent, en conséquence, masquer des altérations présentes chez le locuteur. Le type de test choisi va plus ou moins donner de l'importance aux processus perceptifs descendants. Plus les mécanismes top-down sont impliqués chez l'auditeur, plus on s'écarte de l'évaluation de la performance du locuteur. Dans un cadre clinique, on s'éloigne de la mesure de l'altération en se plaçant sur le versant de l'invalidité, voire de son potentiel handicap au sens de la terminologie de l'OMS. C'est le cas des tests de compréhensibilité qui incluent du décodage acoustico-phonétique (processus ascendant inhérent à tous les tests), de l'accès lexical mais prennent également en compte le contexte de l'échange entre les interactants et tous les autres moyens que le patient met en œuvre pour se faire comprendre (gestes, mimiques, connaissances implicites…). C'est la raison pour laquelle la compréhensibilité reste difficilement quantifiable et qu'on préfère mesurer l'intelligibilité dont on peut obtenir des scores (Woisard et al., 2013).

1.2 Limiter les effets « top-down »

De façon classique, les tests d'intelligibilité sont effectués à partir de phrases ou de mots issus de listes de référence. Les limitations de ce type de test résident dans la capacité des auditeurs à restaurer les séquences distordues. Cet effet est d'autant plus fort que les auditeurs ont une connaissance forte des mots utilisés dans le test et que ces mots sont peu ambigus et donc fortement prédictibles. C'est généralement le cas des orthophonistes qui peuvent faire un usage si important de ces listes qu'ils/elles finissent par les connaître par cœur. On peut citer par exemple la BECD (Auzou et al., 2006) qui ne comporte que 50 mots. Le biais lié à cette connaissance et donc à la forte influence des mécanismes perceptifs descendants est un score d'intelligibilité surévalué car la restauration phonémique de l'auditeur rend « transparentes » les distorsions de production.

2 L'intelligibilité par le biais de 90 000 pseudo-mots

2.1 Construction et principes du test

La solution que nous proposons consiste à utiliser des pseudo-mots, c'est-à-dire des logatomes respectant les structures phonotactiques fréquentes du français, en grande quantité de façon à complètement neutraliser les effets de lexicalité ou d'apprentissage des items par les auditeurs. Au final, les auditeurs sont confrontés à une tâche de décodage acoustico-phonétique suivie d'une transcription écrite. Les détails de la construction du test sont donnés dans (Ghio et al., 2016). Le principe du test est de faire prononcer 52 pseudo-mots tirés aléatoirement d'une liste de 89346 formes possibles, sachant que chaque liste est, par construction, phonétiquement équilibrée. Les pseudo-mots ont été construits avec les formes $C(C)_1V_1C(C)_2V_2$ où $C(C)_i$ est une consonne isolée ou un groupe consonantique. Par exemples: stoumo, vurtant, muja, charou, leba, ranto…

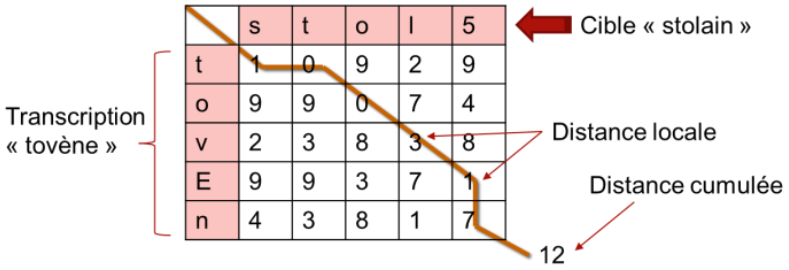
Pour permettre l'énonciation des pseudo-mots par les locuteurs, nous avons utilisé le logiciel PERCEVAL-LANCELOT (www.lpl-aix.fr/~lpldev/perceval/). Le locuteur est placé devant un écran sur lequel est affiché automatiquement le pseudomot à prononcer et une version sonore est produite de façon synchronisée. Cette double modalité, visuelle et auditive, permet de limiter les erreurs de lecture, les limitations auditives et attentionnelles. Etant donnée la taille importante du corpus (89346 formes possibles), les versions sonores sont issues de la synthèse Voxygen (voxygen.fr/). Le locuteur est alors enregistré. Ses enregistrements sont ensuite segmentés semi-automatiquement pour obtenir un fichier audio par logatome produit. L'ensemble des stimuli de tous les locuteurs est finalement soumis à un ensemble d'auditeurs dont la tâche est de transcrire ce qu'ils entendent via le logiciel LANCELOT.

2.2 Le traitement des transcriptions orthographiques

La consigne donnée aux auditeurs pour transcrire orthographiquement chaque pseudo-mot produit par les locuteurs est la suivante : « *Vous allez entendre des non-mots. Un non-mot est une combinaison de sons de la langue française qui n'a pas de signification (ex: gloutu). En respectant les règles de l'orthographe du français, vous devrez transcrire ce que vous entendrez. Certaines prononciations seront difficiles à identifier mais dans tous les cas, vous devrez proposer une transcription.* » Les auditeurs sont choisis natifs de langue française, sans problème auditif et ayant une bonne maitrise de l'orthographe.

Une fois les transcriptions orthographiques recueillies, l'objectif est d'en extraire une forme phonémique car le passage par l'orthographe n'est qu'une étape intermédiaire pour accéder à une représentation phonétique. Les transcriptions orthographiques sont donc phonétisées par l'algorithme LIA_PHON (Bechet, 2001) et elles sont comparées aux formes phonétiques attendues des pseudo-mots. Traditionnellement, par facilité de traitement, le résultat est binaire : correct ou incorrect. Pour dépasser cette évaluation sommaire, nous proposons un résultat analogique sous forme de distance à la cible.

Pour l'opération de comparaison, nous avons utilisé un algorithme de Wagner-Fischer qui intègre les phénomènes d'insertion, élision et substitution d'unités (Figure 1). Dans notre cas, ce calcul de distance de Levenshtein portant non pas sur des unités orthographiques mais sur les phonèmes, il nous est apparu important d'établir une distance locale entre unités (Ghio, 1997). En effet, sur les formes orthographiques, de façon traditionnelle, la distance entre 2 graphèmes est nulle s'ils sont égaux et vaut 1 s'ils sont différents. Dans le cas de phonèmes, il est possible d'apporter des nuances plus subtiles car, par exemple, on peut considérer qu'une confusion entre 2 voyelles n'a pas le même poids qu'entre une voyelle et une consonne sourde.



	s	t	o	l	5
t	1	0	9	2	9
o	9	9	0	7	4
v	2	3	8	3	8
E	9	9	3	7	1
n	4	3	8	1	7

Figure 1 : Comparaison de 2 chaines phonémiques par l'algorithme de Wagner-Fischer (conventions : « 5 » = /ɛ̃/ , « E » = /ɛ/)

3 L'établissement de la matrice de coût entre phonèmes

3.1 La métrique

La matrice de « coût » est un tableau qui contient le degré de dissimilitude entre phonèmes. Elle comporte les 35 phonèmes / a i u o ɔ e ɛ y œ ø ɔ̃ ã ɛ̃ œ̃ p t k b d g f s ʃ v z ʒ m n l R j w ɥ ñ ŋ / auxquels s'ajoutent divers archiphonèmes : Ô = /o/ ou /ɔ/, Ê = /e/ ou /ɛ/, Û = /ø/ ou /œ/, µ= /ɛ̃/ ou /œ̃/, & = /e/ ou /ɛ/ ou /ø/ ou /œ/. Pour le codage des unités phonologiques au format informatique, nous avons utilisé la convention de lexique.org (www.lexique.org/listes/liste_codes_phono.php) car elle a l'avantage de coder une unité sur un caractère, contrairement au codage SAMPA dont la correspondance se fait sur 1 ou 2 caractères, ce qui complique le codage.
Pour constituer la matrice, deux stratégies peuvent être adoptées:

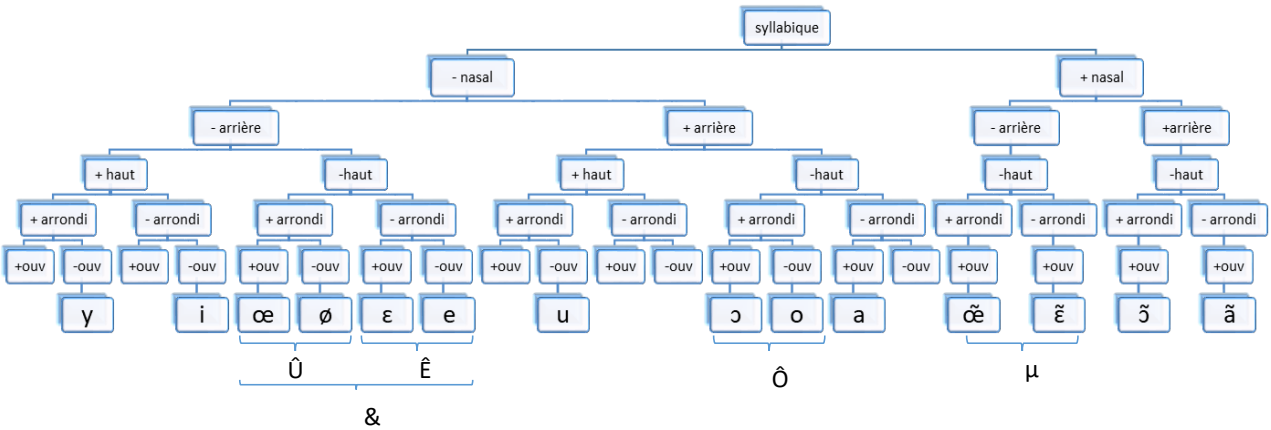
- Une mesure fondée sur les données. Dans ce cas-là, des procédures automatiques calculent statistiquement l'écart moyen entre phonèmes. Il s'agit alors de choisir un corpus représentatif ainsi qu'une métrique pertinente de comparaison.
- Une mesure fondée sur les connaissances. Dans ce cas-là, la distance entre phonèmes est attribuée a priori à partir de savoirs partagés.

Les résultats de nos tests d'intelligibilité pouvant être utilisés comme base d'apprentissage de mesures issus du traitement automatique, nous avons voulu éviter une forme de circularité et avons donc écarté la 1ère solution. Nous avons choisi la seconde méthode.

Afin de réduire son aspect arbitraire, nous avons fondé la comparaison sur la théorie des traits, c'est-à-dire sur le fait que les phonèmes peuvent être décomposés en un ensemble de traits qui les distingue (Jakobson et al., 1951). Il est facile de construire, à partir de cette décomposition, un espace multidimensionnel dans lequel chaque phonème est repéré géométriquement. La notion de traits imposant un caractère binaire (présent ou absent), les coordonnées des phonèmes dans l'espace multidimensionnel ne prennent que les valeurs 0 ou 1. Cela diminue grandement l'importance du choix de la norme. En effet, dans ce cas-là, la valeur donnée par une distance euclidienne $\left(d=\sqrt{\sum_i (x_i-y_i)^2}\right)$ est la racine carrée de celle fournie par une distance de norme 1 $\left(d=\sum_i |x_i-y_i|\right)$. Il n'existe qu'un effet de contraction que nous n'étudierons pas. Nous avons préféré utiliser la distance de norme 1, qui consiste finalement à compter le nombre de traits différents entre deux phonèmes.

3.2 La matrice de coût des voyelles

La Table 1 présente la décomposition en traits distinctifs des voyelles du français d'après Chomsky et Halle (1968). Nous avons remplacé la dénomination chomskyenne [+/- bas] par [+/- ouvert] car moins sujette à confusion avec le trait [+/- haut] qui n'est pas l'opposé du trait [+/- bas]. Dans ce cadre, les voyelles moyennes /e ø o/ sont [-haut ; -bas] et s'opposent respectivement à /ɛ œ ɔ/ qui sont [+bas], c'est-à-dire [+ouvert] dans notre dénomination. La décomposition en arbre permet de mettre en évidence la notion d'archiphonème, c'est-à-dire la sous-spécification d'un trait. Ainsi, les archiphonèmes Ê={e, ɛ}, Û={œ, ø}, Ô={o, ɔ} sont des unités où le trait d'ouverture n'est pas spécifié ; de même, µ={ œ̃, ɛ̃ } et &={ Ê, Û} neutralisent le trait d'arrondissement (labialisation).



	a	i	u	o	e	y	ø	ɛ	ɔ	œ	Ô	Û	Ê	&	ã	ɛ̃	ɔ̃	œ̃	µ
nasal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
arrière	1	0	1	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0
haut	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
arrondi	0	0	1	1	0	1	1	0	1	1	1	1	0		0	0	1	1	
ouvert	1	0	0	0	0	0	0	1	1	1					1	1	1	1	1

Table 1 : Décomposition en traits des voyelles du français sous forme d'arbre et de matrice

Cette décomposition permet ainsi de dresser une matrice de distances entre voyelles par comptage du nombre de traits différents entre chaque phonème (Table 2). La matrice est symétrique.

	a	i	u	o	e	y	ø	ɛ	ɔ	œ	ã	ɛ̃	ɔ̃	œ̃	Ê	Ô	Û	µ	&
a	0	3	3	2	2	4	3	1	1	2	1	2	2	3	1	1	2	2	1
i	3	0	2	3	1	1	2	2	4	3	4	3	5	4	1	3	2	3	1
u	3	2	0	1	3	1	2	4	2	3	4	5	3	4	3	1	2	4	2
o	2	3	1	0	2	2	1	3	1	2	3	4	2	3	2	0	1	3	1
e	2	1	3	2	0	2	1	1	3	2	3	2	4	3	0	2	1	2	0
y	4	1	1	2	2	0	1	3	3	2	5	4	4	3	2	2	1	3	1
ø	3	2	2	1	1	1	0	2	2	1	4	3	3	2	1	1	0	2	0
ɛ	1	2	4	3	1	3	2	0	2	1	2	1	3	2	0	2	1	1	0
ɔ	1	4	2	1	3	3	2	2	0	1	2	3	1	2	2	0	1	2	1
œ	2	3	3	2	2	2	1	1	1	0	3	2	2	1	1	1	0	1	0
ã	1	4	4	3	3	5	4	2	2	3	0	1	1	2	2	2	3	1	2
ɛ̃	2	3	5	4	2	4	3	1	3	2	1	0	2	1	1	3	2	0	1
ɔ̃	2	5	3	2	4	4	3	3	1	2	1	2	0	1	3	1	2	1	2
œ̃	3	4	4	3	3	3	2	2	2	1	2	1	1	0	2	2	1	0	1
Ê	1	1	3	2	0	2	1	0	2	1	2	1	3	2	0	2	1	1	0
Ô	1	3	1	0	2	2	1	2	0	1	2	3	1	2	2	0	1	2	1
Û	2	2	2	1	1	1	0	1	1	0	3	2	2	1	1	1	0	1	0
µ	2	3	4	3	2	3	2	1	2	1	1	0	1	0	1	2	1	0	1
&	1	1	2	1	0	1	0	0	1	0	2	1	2	1	0	1	0	1	0

Table 2 : matrice de coût des voyelles (⇔ nombre de traits différents entre les voyelles)

3.3 La matrice de coût des consonnes

Dans la décomposition des consonnes du français, un certain nombre de traits est clairement défini :

- Le trait vocalique (+/- sonant) distingue les obstruantes (occlusives et fricatives : -sonant) des consonnes liquides (l R), nasales (m n ñ) et semi-voyelles (j w ɥ) : +sonant
- Le trait de nasalité distingue les consonnes nasales (+nasal) des orales (-nasal)
- Le trait de voisement distingue les consonnes sonores (voisées) des sourdes (-vois)
- Le trait de continuité distingue les occlusives (-cont) des fricatives (+cont).
  Parmi les consonnes vocaliques, Chomsky et Halle (1968) précisent p.317 que les occlusives nasales sont considérées comme interrompues (-cont). Les auteurs précisent enfin que le cas de /l/ et /r/ est complexe mais finissent par proposer un trait (+cont) à /r/ et (-cont) à /l/. Cette caractérisation est confirmée dans Clements (2005) p.47.

En revanche, les traits relatifs au lieu d'articulation de la consonne posent de multiples problèmes. En effet, d'après l'alphabet phonétique international (www.internationalphoneticalphabet.org), les consonnes du français sont articulées selon 7 lieux différents qui peuvent être regroupés en 3 grandes classes d'articulation : les labiales, les dentales et les vélo-palatales (Table 3).

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar
Plosive	p b			t d			k g
Nasal	m			n		ɲ	ŋ
Fricative		f v		s z	ʃ ʒ		
approximant				l		j	

Labiales Dentales Vélo-Palatales

Table 3 : lieu d'articulation des consonnes du français (d'après l'IPA)

Dans une approche totalement phonologique, Chomsky et Halle (1968) proposent p.223 la décomposition selon les deux traits +/- coronal (pointe de la langue) et +/- antérieur, ce qui donne la Table 4 ci-dessous. Cette décomposition place alors /p/ (-cor, +ant) à un trait de distance de /t/ (+cor) et à un trait de distance de /k/ (-ant). En revanche, il place /t/ (+cor, +ant) à deux traits d'écart de /k/ (-cor, -ant), ce qui n'est pas très satisfaisant d'un point de vue articulatoire où il semblerait logique de respecter l'ordre /p t k/, c'est-à-dire /t/ équidistant de /p/ et /k/, /p/ et /k/ étant plus éloignés.

	+ coronal	-coronal
+antérieur	Dental : t d s z	Labial : p b f v
-antérieur	Palato-alveolaire : rien en français	Vélaire : k g (ʃ ʒ)

Table 4 : traits relatifs au lieu d'articulation d'après Chomsky et Halle (1968)

Clements (2005) propose une décomposition en 3 traits exclusifs : labial, coronal et dorsal qui reflètent directement les 3 lieux décrits en Table 3. Nous estimons qu'il y a là une sur spécification car 2 traits seulement sont nécessaires pour coder 3 états. Nous avons finalement opté pour les travaux de Jakobson et al. (1951) qui proposent 2 traits acoustiques permettant une distinction adéquate :

- Le trait compact/diffus : "the consonants articulated against the hard or soft palate (velars and palatals) are more compact than the consonants articulated in the front part of the mouth." (Jakobson et al., 1951, p.27)
- Le trait grave/aigu : " gravity characterizes labial consonants as against dentals, as well as velars vs. palatals" (Jakobson et al., 1951, p.30)

Nous obtenons finalement la décomposition des consonnes en traits (Table 5) et la matrice de distances entre consonnes (Table 6).

	p	t	k	b	d	g	f	s	S	v	z	Z	m	n	N	l	R	j	w	ɥ
vocalique	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
continu	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1
nasal	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
voisé	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1
compact	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
aigu	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	1	1	1	0	0

Table 5 : Décomposition en traits des consonnes du français

	p	t	k	b	d	g	f	s	S	v	z	Z	m	n	N	l	R
p	0	1	2	1	2	3	1	2	3	2	3	4	3	4	5	3	4
t	1	0	1	2	1	2	2	1	2	3	2	3	4	3	4	2	3
k	2	1	0	3	2	1	3	2	1	4	3	2	5	4	3	3	4
b	1	2	3	0	1	2	2	3	4	1	2	3	2	3	4	2	3
d	2	1	2	1	0	1	3	2	3	2	1	2	3	2	3	1	2
g	3	2	1	2	1	0	4	3	2	3	2	1	4	3	2	2	3
f	1	2	3	2	3	4	0	1	2	1	2	3	4	5	6	4	3
s	2	1	2	3	2	3	1	0	1	2	1	2	5	4	5	3	2
S	3	2	1	4	3	2	2	1	0	3	2	1	6	5	4	4	3
v	2	3	4	1	2	3	1	2	3	0	1	2	3	4	5	3	2
z	3	2	3	2	1	2	2	1	2	1	0	1	4	3	4	2	1
Z	4	3	2	3	2	1	3	2	1	2	1	0	5	4	3	3	2
m	3	4	5	2	3	4	4	5	6	3	4	5	0	1	2	2	3
n	4	3	4	3	2	3	5	4	5	4	3	4	1	0	1	1	2
N	5	4	3	4	3	2	6	5	4	5	4	3	2	1	0	2	3
l	3	2	3	2	1	2	4	3	4	3	2	3	2	1	2	0	1
R	4	3	4	3	2	3	3	2	3	2	1	2	3	2	3	1	0

Table 6 : matrice de coût des consonnes (⇔ nombre de traits différents entre les consonnes)

3.4 Les distances inter macro-classes

Les semi-consonnes /j w ɥ/ ont été placées de façon identique à leur équivalent /i u y/ mais avec le trait de syllabicité en moins (-syll). En effet, ces phonèmes ne peuvent à eux seuls constituer une syllabe (Chomsky & Halle, 1968). Dans leurs distances aux consonnes, elles ont été décomposées comme présentées en Table 5.

Par rapport aux voyelles, les consonnes ont été placées à une distance supérieure à la distance maximale entre voyelles (d=6). En tenant compte de la classification de Dell (1985),

<table>
<tr><td rowspan="2">non syllabique</td><td rowspan="2">consonantique</td><td rowspan="2">non vocalique</td><td>sourd</td><td>Obstruantes sourdes</td></tr>
<tr><td rowspan="3">sonore</td><td>Obstruantes sonores</td></tr>
<tr><td rowspan="2">non consonantique</td><td rowspan="2">vocalique</td><td>consonnes nasales et liquides</td></tr>
<tr><td>syllabique</td><td>semi-voyelles</td></tr>
<tr><td></td><td></td><td></td><td></td><td>voyelles</td></tr>
</table>

nous avons ensuite respecté la hiérarchie suivante :

Voyelles < Liquides < Nasales < Obstruantes sonores < Sourdes

Au final, nous obtenons une matrice de « coût » qui contient le degré de dissimilitude entre les 35 phonèmes retenus pour le français.

4 Application : La mesure d'intelligibilité de patients avec traitement du cancer des voies aériennes supérieures

Le protocole décrit précédemment a été utilisé dans le cadre du projet C2SI (Carcinologic Speech Severity Index) dont l'objectif est d'obtenir une mesure de l'impact des traitements des cancers de la cavité buccale et du pharynx sur la production de la parole par l'Indice de sévérité des troubles de la production de la parole à la fois par des méthodes perceptives et par traitement automatique de la parole.

126 locuteurs (41 sujets sains et 85 patients) enregistrés dans le service d'oncoréhabilation de l'Oncopole à Toulouse ont produit chacun 52 pseudo-mots tirés aléatoirement de la liste de 89346 formes possibles. 40 auditeurs ont retranscrit ces productions, chaque pseudo-mot d'un locuteur étant transcrit par 3 auditeurs différents. Ces tests se sont déroulés au sein du Centre d'Expérimentation sur la Parole (www.lpl-aix.fr/~cep) du Laboratoire Parole et Langage à Aix-en-Provence. Les transcriptions orthographiques ont été phonétisées et comparées aux formes phonétiques attendues des pseudo-mots par l'algorithme décrit au §2.2

De façon globale, les résultats montrent que les formes perçues chez les sujets sains sont en moyenne à une distance de 0.48 trait/phonème (sdev= 0.22) par rapport aux formes attendues alors que cette distance passe à 1.28 (sdev=0.63) pour les patients. En effectuant une transformation logarithmique du score, les distributions deviennent normales (test de Shapiro ; $p > 0.05$) et les variances homogènes (test de Bartlett ; $p > 0.05$). La différence entre les deux groupes est significative ($p < 0.01$).

Dans l'avenir, nous allons nous employer à vérifier l'équivalence des listes, c'est-à-dire que nous allons mesurer quels sont les écarts obtenus sur un même locuteur produisant plusieurs listes. De plus, ces mesures vont être comparées à des évaluations cliniques globales ainsi qu'à des mesures acoustiques automatiques (Astésano et al. , 2018).

5 Conclusion

Nous avons mis au point un test d'intelligibilité à partir d'une importante cohorte de pseudo-mots répondant aux contraintes phonotactiques du français. Cette méthode a été testée sur 126 locuteurs en milieu hospitalier sans obstacle majeur. La transcription orthographique par des auditeurs est suivie d'une transformation graphème-phonème puis d'une comparaison sophistiquée à la cible phonétique attendue. Cette comparaison fondée sur un calcul de traits distinctifs a l'immense avantage de créer une métrique progressive différente des approches traditionnelles qui se contentent de compter le nombre d'occurrences correctes. La construction même du test ne permet aucune restauration phonémique par les auditeurs des séquences mal produites par les locuteurs. Il n'y a donc pas d'effet plafond, ce qui pourra permettre de quantifier finement par exemple des effets thérapeutiques. Pour conclure, le test semble discriminant en ce qui concerne la mesure de la performance articulatoire des locuteurs.

Remerciements

Ce travail fait partie du projet C2SI (Carcinologic Speech Severity Index) financé par l'Institut National du Cancer dans le cadre de projets libres de recherche en Sciences Humaines et Sociales, Epidémiologie et Santé Publique. L'investigatrice principale est Virginie Woisard du CHU Larrey à Toulouse. Nous remercions la Sté Voxygen pour avoir synthétisé les 89346 stimuli du corpus. Nous remercions le personnel du CEP (www.lpl-aix.fr/~cep), en particulier Carine André, pour la réalisation des tests de perception.

Références

ASTESANO C. , BALAGUER M., FARINAS J., FREDOUILLE C., GAILLARD P., GHIO A., GIUSTI L. et al. (2018), Carcinologic Speech Severity Index Project: A Database of Speech Disorders Productions to Assess Quality of Life Related to Speech After Cancer, LREC, 7-12 May 2018, Miyazaki (Japan)

AUZOU P, ROLLAND-MONNOURY V. (2006), Batterie d'évaluation de la dysarthrie, *1st ed. Isbergues: Ortho Edition*.

BECHET F (2001), LIA_PHON : UN SYSTEME COMPLET DE PHONETISATION DE TEXTES, TRAITEMENT AUTOMATIQUE DES LANGUES - TAL - VOLUME 42 NUMERO 1 - PP 47-67, 2001

CLEMENTS G.N. (2005), The role of features in speech sound inventories In Raimy & Cairns, eds., *Contemporary Views on Architecture and Representations in Phonological Theory*. Cambridge, MA: MIT Press, p 19-68

CHOMSKY N., HALLE M. (1968), *The Sound Pattern of English*. New York: Harper & Row.

DELL F. (1985), LES REGLES ET LES SONS, HERMANN, PARIS.

GANONG W. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology. *Human perception and performance*, 6 (1), 110-125.

GHIO A., GIUSTI L., BLANC E., PINTO S., LALAIN M, ROBERT D., FREDOUILLE C., WOISARD V. (2016) Quels tests d'intelligibilité pour évaluer les troubles de production de la parole ?. *Journées d'Etude sur la Parole*, Paris, France, p.589-596

GHIO A.(1997) Achile : un dispositif de décodage acoustico-phonétique et d'identification lexicale indépendant du locuteur à partir de modules mixtes. Thèse de l'Université d'Aix Marseille, 1997.

JAKOBSON R., FANT G., HALLE M. (1951), "Preliminaries to speech analysis", MIT Press, Cambridge.

WARREN RM., WARREN RP. (1970), Auditory illusions and confusions. *Sci. Am.*; 223, 30-36

WOISARD V., ESPESSER R., GHIO A., DUEZ D. (2013). De l'intelligibilité à la compréhensibilité de la parole, quelles mesures en pratique clinique ? Revue de laryngologie, otologie, rhinologie, vol. 1, no. 134. 2013, p. 27-33.