

Introduction

Les recherches antérieures ont montré que la complexité phonétique a une influence sur le maintien de la production de la parole fluide (Coalson et al., 2012, Howell, 2004).

Contrôler cette variable est ainsi important dans les évaluations langagières ou de la mémoire utilisant des stimuli verbaux.

A partir de la théorie « cadre puis contenu » de MacNeilage et Davis (1990), selon laquelle le développement vocal précoce est conditionné par la maturation physiologique et la maîtrise des habiletés motrices de l'appareil articulatoire, Jakielski (2000) a développé l'Index of Phonetic Complexity (IPC) qui permet de quantifier la complexité phonétique d'un mot à travers 8 paramètres.

IPC a été utilisé dans divers domaines d'étude comme l'acquisition du langage (Charlier Bererd et al., 2007), la pathologie du langage telle que la dyslexie (Bose et al., 2011), le bégaiement (Lasalle et al., 2011), l'autisme (Ferré et al., 2013), la maladie d'Alzheimer (Lee, 2012).

Cet indice initialement conçu pour l'anglais a été utilisé dans différentes langues, par exemple en allemand (Dworzynski et al., 2004) et en espagnol (Howell et al., 2007).

Notre objectif est de développer un outil de calcul automatique de la complexité phonétique qui fournit un résultat rapide et objectif.

Méthode

N°	Paramètres		Exemples en français
1	Lieu d'articulation de la consonne	Dorsales	[k], [g], [x], [ŋ] « agrafeuse » [aɡʁaføz] = 2 points
2	Mode d'articulation de la consonne	Fricatives, affriquées, liquides	[f], [v], [s], [z], [ʃ], [ʒ], [ʎ], [ʝ] « calculatrice » [kalkylatʁis] = 4 points
3	Classe de voyelle	Rhotiques	Ce paramètre n'étant pas pertinent pour le français, il a été écarté dans cette étude
4	Final du mot	Consonne	« ark » [aʁk] = 1 point
5	Longueur de mot	≥ 3 syllabes	« bibliothèque » [bi-bli-jø-tɛk] = 1 point
6	Variation de lieu d'articulation des consonnes isolées	Variée	« baguette » [bagnet] = 1 point
7	Groupe consonantique	Présence	« rétroprojecteur » [ʁɛtʁopʁɔʒɛktœʁ] = 3 points
8	Type de groupe consonantique	Hétéro-organique	« brique » [bʁik] = 1 point

Tableau - Index of Phonetic Complexity (Jakielski, 2000)

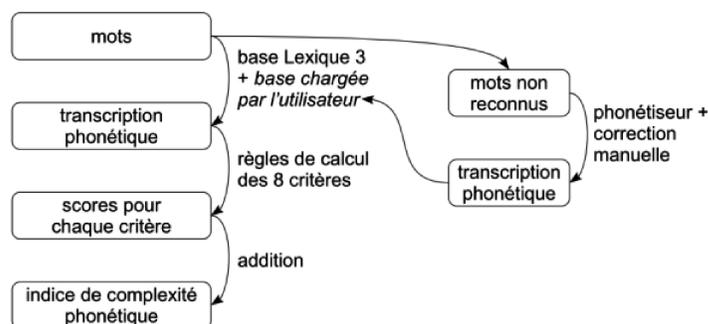


Figure 1 - Schéma descriptif de l'approche automatique implémentée dans iPhocomp



Index of Phonetic Complexity

Résultats obtenus

2 transcriptions personnalisées prises en compte dans le dictionnaire phonétique.

[Cacher les résultats détaillés](#)

Mot	Prononciation	Param. 1	Param. 2	Param. 3	Param. 4	Param. 5	Param. 6	Param. 7	Param. 8	IPC
pêche	pES	0	1	0	1	0	1	0	0	3
agenda	aZ5da	0	1	0	0	1	0	0	0	2
samedi	sam'di	0	1	0	0	1	1	0	0	3
vis	vis	0	2	0	1	0	1	0	0	4
jep		0	0	0	0	0	0	0	0	0
TOTAL		0	1.25	0	0.5	0.5	0.75	0	0	3

Param. 1	Param. 2	Param. 3	Param. 4	Param. 5	Param. 6	Param. 7	Param. 8	IPC
0	1.25	0	0.5	0.5	0.75	0	0	3

Attention, la liste de mots analysés contient 3 mots potentiellement problématiques : un mot non reconnu donc pas analysé, un mot avec de possibles homographes non homophones, un problème de e caduc.

Temps passé : 0.299117 ms.

4 mots reconnus / 5 : 80%.

Mots non reconnus

Les formes en minuscules sont données, la forme originale est indiquée entre parenthèses :

- jep (JEP)

Transcription des mots non reconnus

Vous pouvez, si vous le souhaitez, enregistrer une transcription phonétique des mots non reconnus pour la charger ensuite lors de l'envoi du texte, afin qu'elle soit prise en compte. Une tentative de transcription automatique basique, qui nécessite une relecture attentive, est donnée sur chaque ligne.

jep:Z'p

Figure 2 - Interface d'iPhocomp

Résultats

• Outil objectif

Une étude comparative entre l'analyse manuelle et l'analyse automatique sur la base de données de Bonin et al. (2003), qui comporte un ensemble de 299 images associées à une dénomination de référence en français, a permis de repérer des erreurs dans l'analyse manuelle (0.4 % d'erreur de type inattention).

iPhocomp permet ainsi d'éviter l'erreur humaine, optimisant la comparabilité des résultats de différentes études.

• Outil rapide

Le temps de calcul humain sur 236 mots, issus du corpus du discours oral spontané des patients atteints de la maladie d'Alzheimer et des sujets âgés sains, est d'environ 2 h 30.

Alors que le temps de calcul automatique sur le même corpus est de 0.54 ms.

Ainsi, iPhocomp permet de réduire significativement le temps total d'évaluation de l'indice de complexité phonétique.

• Diverses possibilités d'analyse

L'utilisateur peut charger une liste personnalisée de transcription phonétique afin d'inclure des mots absents de base de données, pour des études sur les pseudo-mots, ainsi que sur les éventuelles variations de prononciation due à des facteurs géographiques, stylistiques ou pathologiques.

iPhocomp fournit la valeur de complexité phonétique des mots, mais aussi les résultats détaillés pour chacun de ces 8 paramètres afin d'ouvrir des perspectives d'analyses plus qualitatives.

iPhocomp peut également calculer la valeur moyenne d'une liste de mots. Cela permet de mener des analyses comparatives en évaluant ces paramètres sur deux listes de mots distinctes.

Conclusion

La nécessité d'approfondir l'étude de la complexité phonétique de la langue française.

L'étude de divers phénomènes observables dans l'acquisition du langage, l'apprentissage de langues étrangères, la pathologie du langage, etc., permettrait de dégager des éléments pertinents pour enrichir notre connaissance, permettant de mieux adapter l'indice de complexité phonétique de Jakielski aux spécificités de la langue française.

Le logiciel iPhocomp est librement mis à la disposition de la communauté scientifique (<http://igm.univ-mlv.fr/~gambette/iPhocomp/>).

Nous espérons que ce logiciel donnera un nouvel élan à la fois aux recherches fondamentales et aux études appliquées.