

La reconnaissance des mots dans la parole accentuée : Une étude en laboratoire et à l'extérieur.

Delphine Dei¹ Page Piccinini²
Isabelle Dautriche¹ Marieke Van Heugten³
Alejandrina Cristia¹

(1) Laboratoire des Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS)
Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University,
29 Rue d'Ulm, 75005, Paris, France

(2) Phonetic Lab, Department of Linguistics, UC San Diego,
9500 Gilman Drive, La Jolla, CA, United States

(3) Department of Psychology, University at Buffalo, State University of New York,
Buffalo, NY, United States

delphine.dei04@gmail.com, page.piccinini@gmail.com,
isabelle.dautriche@gmail.com, mariekev@buffalo.edu,
alecristia@gmail.com

RESUME

Des travaux récents suggèrent que les enfants et les adultes sont initialement ralentis dans leur compréhension des mots qui n'ont pas été prononcés de façon standard. Néanmoins, quand ils font face à un interlocuteur qui à un discours accentué, ils développent rapidement des stratégies spécifiques qui leur permettent de comprendre même des prononciations atypiques. Cependant, ces résultats sont typiquement issus de recherches en laboratoire, où l'attention des participants se concentre sur une tâche unique qui leur demande peu de ressources. Afin de dépasser ces limitations, nous avons mené une expérience de reconnaissance de mots sur tablette tactile, en évaluant des enfants et des adultes, en laboratoire et dans l'environnement naturel de chaque groupe. Nous avons constaté que des déviations de prononciation dans la parole accentuée ralentissent la reconnaissance des mots, chez des enfants et adultes, tant dans le laboratoire que dans des environnements naturels.

ABSTRACT

Mispronunciations slow down word recognition: A study using touchscreens in the lab and the real world.

Recent work suggests that children and adults are initially delayed when exposed to an accented talker, but quickly recover and develop talker-specific listening patterns. This research, however, is often carried out in the lab, where participants' attention is focused on a single task and no distractions are present, and recovery is often deduced from looking patterns that require little overt response. In this experiment, we overcome these limitations and assess children's and adults' word recognition in accented speech, tested in both a laboratory and in a naturalistic setting. We found that mispronunciations in accented speech slow down overt decisions about word recognition, to a similar extent in the lab and in familiar environments, for both children and adults.

MOTS-CLES : reconnaissance de mots, parole accentuée, prononciations modifiées, acquisition du langage, enfance.

KEYWORDS: word recognition, accented speech, mispronunciations, language acquisition toddlerhood.

1 Introduction

Bien que les accents étrangers et régionaux puissent varier à tous les niveaux phonétiques et linguistiques, nous nous adaptons rapidement aux accents auxquels nous ne sommes pas habitués. Afin de mieux comprendre le système cognitif qui permet la compréhension de la parole, de nombreux travaux ont été menés sur la perception des mots dans la parole accentuée. Ces études ont été conduites principalement chez des adultes, et plus récemment chez des enfants.

Plusieurs expériences en laboratoire suggèrent une flexibilité remarquable qui permet aux enfants et aux adultes de s'habituer rapidement à un nouvel accent, et à accepter des prononciations différentes de celles attendues (voir Cristia et al., 2012, pour un résumé sur la perception de la parole accentuée chez les enfants, jeunes adultes, et des populations plus âgées). Par exemple, Schmale, Cristia, et Seidl (2015) ont montré à des enfants de deux ans des vidéos de personnes physiquement très différentes : un enfant, une dame âgée, une jeune fille, etc. Après cette exposition purement non linguistique, les enfants pouvaient reconnaître des mots nouvellement appris dans un accent étrange qu'ils n'avaient pas entendu auparavant. D'autres études de laboratoire coïncident avec cette vision à peine quelques minutes d'exposition suffisent pour commencer à accepter des prononciations déviantes (par ex., Maye, Aslin, & Tanenhaus, 2010; Van Heugten & Johnson, 2014).

Or, d'autres études suggèrent que les adultes et les enfants sont relativement conservateurs, en ne gardant dans leur lexique que des caractéristiques typiques de leur communauté. Par exemple Girard, Floccia, et Goslin (2006) ont étudié la perception de mots prononcés dans des accents régionaux typiques de Besançon ou Toulouse auprès d'adultes résidants dans les deux régions. Ils ont trouvé que la familiarité avec l'accent prédisait la performance dans une tâche de décision lexicale, suggérant que l'expérience est cruciale pour la compréhension de la parole accentuée. Les résultats de Floccia et al. (2012) sont encore plus remarquables: les enfants de 20 mois britanniques seraient *incapables* de reconnaître un mot dans la prononciation utilisée par leurs parents si cette prononciation n'est pas celle de la communauté où l'enfant grandit. Il faut cependant indiquer que ces résultats surprenants n'ont pas été répliqués lors d'une étude plus récente menée aux Pays-Bas (Van der Feest & Johnson, 2016).

Globalement, nous voyons que la littérature ne montre pas une image monolithique et simple de la perception de la parole accentuée. Il est possible que les écarts de résultats soient, au moins en partie, dus à des différences dans la complexité de la tâche, dans la façon dans laquelle elle a été conçue, et dans la prévalence des accents dans le lieu où elle a été menée. Une deuxième limitation des études antérieures est qu'elles ont été conduites dans des lieux inhabituels tels que les laboratoires, bien loin des conditions du monde réel. En effet en dehors du laboratoire nous sommes bien souvent entourés d'autres individus, et probablement plus ouverts à la possibilité de rencontrer des accents et individus variés. Cette différence pourrait se révéler capitale car nos modèles de la perception de la parole sont donc basés sur des expériences peu écologiques qui pourraient ne pas être généralisables à la perception de la parole comme elle se déroule dans la vie de tous les jours. Or, du moins à notre connaissance, la possibilité que le contexte expérimental joue un rôle dans la flexibilité face à des prononciations peu standards n'a pas été étudiée.

L'expérience que nous proposons aborde ces limitations et compare en laboratoire et dans un lieu quotidien la reconnaissance de mots accentués. Des travaux précédents ont montré que la perception de la parole accentuée est facilitée par des indices extérieurs, tels que la présence d'une photo révélant l'identité du locuteur (Johnson, Strand, & D'Imperio, 1999) ou même la présence d'un mot ou d'un signe régional (Hay & Drager, 2010). Notre étude ne cherche pas à étudier ces biais, mais plutôt à déterminer si les participants qui entendent une parole accentuée sont plus flexibles, c'est-à-dire, acceptent des prononciations déviantes plus facilement, quand ils sont testés dans un environnement quotidien, par rapport au laboratoire. Nous avons donc recruté des adultes et de jeunes enfants, car l'impact de l'environnement pourrait être plus marqué chez ces derniers.

Tous les participants ont réalisé une tâche de choix forcé avec des stimuli enregistrés auprès d'une personne ayant un accent étranger (anglais britannique, un accent commun à Paris). Les participants avaient en main une tablette tactile où 2 images à la fois étaient visibles (par ex. une maison, une chaussure) et où un personnage animé leur demandait d'appuyer sur l'une d'elle (ex "Touche la maison"). Les objets visibles n'avaient aucun recouplement phonologique de sorte que le choix était toujours évident. Or, dans certains essais, le mot n'était pas prononcé de la façon attendue (par ex. "bateau" devenait "pateau" avec l'accentuation anglaise). Nous voudrions savoir si les enfants et les adultes sont moins affectés par ces déviations dans un environnement quotidien par rapport à la même tâche en laboratoire. Puisque notre recrutement se passait en partie dans des lieux publics, il a été impossible de faire des groupes *a priori* sur la base de l'expérience linguistique des participants. Néanmoins, un questionnaire nous a permis de contrôler ces caractéristiques post hoc.

2 Méthode

2.1 Participants

Nous présentons les caractéristiques des participants séparés par groupe d'âge (adultes, enfants) et du lieu où le test s'est déroulé (laboratoire, quotidien), comme illustré dans la Table 1. Le premier groupe d'enfants a été testé au laboratoire bébé du LSCP (groupe « Enfant-Labo »). Le deuxième groupe a été testé à la crèche l'Arbre-Sec (Paris 1er) pour les plus jeunes et à l'école François Coppée (Paris 15e) en petite section pour les plus grands (groupe « Enfant-Quotidien »). De la même façon, nous avons testé une partie des adultes en cabine de test au laboratoire (« Adulte Labo »), et d'autres à la bibliothèque universitaire de l'UPMC (« Adulte-Quotidien »). Dans tous les cas, les participants ou leurs représentants légaux ont donné leur consentement par écrit.

À l'intérieur de ces quatre groupes, des participants ont dû être exclus en suivant un critère d'exposition au français non-accentué minimum de 80% du temps, ainsi que 7 critères d'exclusion fixés avant l'inspection des résultats : 1) Le questionnaire d'expérience linguistique n'a pas été rempli ; 2) Le français est entendu moins de 50% du temps ; 3) Réponses incorrectes aux deux essais de familiarisation ; 4) Plus de 25% de réponses incorrectes aux essais correctes pour un francophone ; 5) Avoir répondu à moins de 9 essais test sur les 12 ; 6) Avoir subi une ou plusieurs otites pendant la semaine qui précède le test ; 7) Avoir été diagnostiqué avec un trouble du développement (hors prématurité).

	Total → Inclus	Âge (écart type)
Enfant-Labo	19 (9) → 9	42,25 (± 8,13) mois
Enfant-Quotidien	48 (13) → 18	36,80 (± 7,45) mois
Adulte-Labo	15 (10) → 8	23,17 (± 4,37) ans
Adulte-Quotidien	28 (15) → 11	22,77 (± 3,78) ans

TABLE 1 : Informations démographiques des groupes testés. La colonne “Total” reporte le nombre total d’individus testés pour chaque groupe (et le nombre de femme parmi eux). “Inclus” indique le nombre d’individus inclus dans nos analyses. “Âge” indique l’âge moyen (et l’écart type) pour les individus qui ont pu être inclus

2.2 Matériel et procédure

Nous avons mesuré le temps de réaction avec une tâche de choix forcé implémentée sur une tablette tactile numérique iPad®. Pour rendre l’activité plus ludique et attractive pour les enfants, les énoncés étaient adaptés à leur âge et formulés par un personnage animé dans un décor coloré (voir Figure 1) Dans chaque essai une paire d’objets était visible et l’un des deux était nommé par le personnage dans la phrase « Touche le ___ ». Les essais n’étaient pas répétés, même en cas de réponse incorrecte. En revanche, le personnage sur l’écran renvoyait un feedback positif en cas de réponse correcte (« Super ! », en sautant de joie) et négatif en cas de réponse incorrecte (« Non, c’est pas celui la », tristement). L’expérimentateur donnait aux enfants des encouragements neutres à la tâche (« tu te débrouilles très bien ») afin de maintenir son intérêt.



FIGURE 1: Capture d’écran de l’iPad® pendant le jeu pour un essai avec le couple banane/chaussette

Il y avait au total 14 essais. Les deux premiers ont servi à familiariser les participants avec la tâche et au fait que le personnage avait un accent étranger. Uniquement durant cette phase les individus testés pouvaient recevoir du feedback de l’expérimentateur sur leur performance. Nous avons donc utilisé des stimuli neutres à notre question de recherche pour cette première phase, les réponses n’étaient prises en compte que dans les critères d’exclusion définis plus haut.

Les 12 essais restants étaient répartis en 3 conditions différentes : **Prononciation française (PF)**, le mot est prononcé avec la forme phonologique attendue avec une prononciation française standard **Prononciation anglaise (PA)**, le mot est prononcé avec la forme phonologique attendue avec un

prononciation accentuée anglaise, par exemple comme vu plus haut: “bateau” devient “*p*ateau” ; et **Prononciation composée (PC)**, le mot est prononcé d’une façon inattendue pour l’accentuation française et anglaise, par exemple le mot “*b*anane” devient “*van*ane”. Les mots test utilisés dans ces trois conditions avaient soit des voyelles nasales, soit des voyelles antérieures arrondies, soit des consonnes occlusives voisées (voir section Stimuli). Ces associations étaient contrebalancées à travers les participants ; par exemple, pour un participant donné tous les mots ayant des voyelles nasales étaient prononcés correctement et donc représentaient la condition PF ; les voyelles de tous les mots avec des voyelles antérieures arrondies changeaient de hauteur, représentant ainsi la condition PC ; et toutes les consonnes occlusives voisées des mots restants étaient dévoisées pour représenter la condition PA.

2.3 Stimuli

Les stimuli ont été produits par une femme qui a pour langue maternelle l’anglais britannique et qui parle français couramment. Nous les avons traité sous PRAAT (Boersma & Weenink, 2015), en nous assurant qu’ils ne différaient ni en intensité ni en durée à travers les 3 conditions (en moyenne).

Les stimuli sélectionnés étaient des noms connus par plus de 60% d’enfants d’après un test de vocabulaire donné à 81 enfants de 18 mois lors d’une étude précédente. Nous avons ensuite sélectionné 14 paires de mots [cibles + distracteurs], 2 pour les essais de familiarisation et 12 pour les essais test. Nous avons exclu les mots qui contenaient plus d’un type de contraste; comme par exemple le mot “ballon” qui possède une consonne occlusive voisée, /b/, et une voyelle nasale, /ɔ̃/. Par ailleurs, nous nous sommes assurés que les prononciations modifiées pour les conditions PA et PC, de tous les mots cibles, ne formaient pas de vrais mots. Les mots prononcés pour la phase de familiarisation et les phrases de transition (feedback par exemple) ne contenaient pas de contraste affectés par l’accent anglais.

Les distracteurs étaient choisis pour avoir le même genre grammatical, nombre de syllabes, et la même animéité que la cible. De plus, les mots cibles et leurs distracteurs appariés étaient clairement phonologiquement différents, ainsi, même si la prononciation était différente de celle attendue, la seule réponse possible restait l’image cible. Les paires cible-distracteur choisies étaient les suivantes : pour les voyelles nasales *poisson-lapin*, *maison-chaussure*, *manteau-gâteau*, *compote-poussette* ; pour les voyelles arrondies *lunettes-pantoufles*, *voiture-poubelle*, *fleur-bouche*, *yeux-chiens* ; pour les consonnes voisées *balle-main*, *bateau-cochon*, *bouteille-cuillère*, *banane-chaussette*.

2.4 Analyses et prédictions

Notre question de recherche principale concerne la flexibilité en perception face à des prononciations non-standards à travers différents environnements de test. Le temps de réponse est défini comme le temps écoulé entre la fin de l’énoncé (ex. Touche la maison) et le moment du touché sur la tablette. Afin d’obtenir une meilleure appréciation des résultats en éliminant les variations individuelles, les différences d’âge, et les variations d’ordre général sur le temps de réponse, nous avons choisi de calculer les ratios des temps de réponse sous R (R Core Development Team, 2008) tel que:

$$([PA] / [PA] + [PF]) - 0,5 \text{ et } ([PC] / [PC] + [PF]) - 0,5$$

Où [PA] correspond à la médiane des temps de réponse pour la prononciation anglaise, [PF] à la médiane des temps de réponse pour la prononciation française et [PC] celle pour la prononciation composée.

Sur la base des travaux antérieurs, et compte tenu du fait que le test était très court, nous nous attendons à ce que les ratios soient supérieurs à 0 au laboratoire, indiquant que les participants sont ralentis par les prononciations inattendues (PA, PC) par rapport à la prononciation attendue (PF). Ceci répliquerait des résultats antérieurs montrant que les enfants et adultes sont sensibles à la prononciation d'un mot même lorsqu'il est prononcé avec un accent étranger. Notre question principale porte sur la flexibilité en perception dans un environnement quotidien par rapport au laboratoire. Si les participants sont plus flexibles dans un environnement quotidien, alors leurs ratios devraient être significativement plus bas dans cet environnement par rapport au laboratoire. Par contre, si le contexte ne fait aucune différence, le ratio devrait être aussi élevé (et supérieur à 0 pour chaque groupe d'âge dans chaque milieu. Pour comparer les ratios, nous utiliserons un test non paramétrique de Wilcoxon entre les deux sous-groupes (environnement quotidien versus laboratoire) à l'intérieur de chaque groupe d'âge.

3 Résultats

Par souci de simplicité, nous présentons seulement les ratios pour la condition PC, mais le pattern de résultats dont nous concluons est le même pour la condition PA. Toutes nos données et nos analyses sont disponibles sur la plateforme Open Science Framework (OSF), et directement accessibles depuis <https://osf.io/pnhc3/>.

Tout d'abord, nous observons que les enfants sont en général plus lents que les adultes, et que les enfants testés au laboratoire sont plus lents que ceux testés dans leur environnement quotidien. Enfants-Labo(EL)=3,93s, EQ=3,47s, AL=2,26s et Adultes-Quotidien (AQ)=2,35s en moyenne. Tous les ratios sont significativement différents de 0 pour tous les groupes (voir Figure 2). Cela signifie que, comme prédit, les temps de réponse pendant des essais où les mots sont prononcés de manière inattendue (prononciation anglaise ou composée) sont supérieurs à ceux des mots prononcés en français standard. Enfin, cet effet n'est pas modulé par le lieu d'expérimentation.

Des analyses supplémentaires ont été faites mais n'ont pas été présentées ici par manque d'espace. Elles peuvent être reproduites en utilisant les données et le code disponible sur notre site OSF. Notamment, nous avons évalué la possibilité qu'un petit nombre de participants ayant une expérience avec d'autres langues ou accents supérieure à 20% de leur temps soient plus flexibles avec les prononciations inattendues. Contrairement à cette hypothèse, nous ne trouvons pas de différence entre les deux groupes de participants (avec et sans expérience). Nous ne trouvons pas non plus de différences marquées parmi les enfants en fonction de leur exposition préalable aux écrans tactiles, ni à travers les différents types de changements (voyelles versus consonnes).

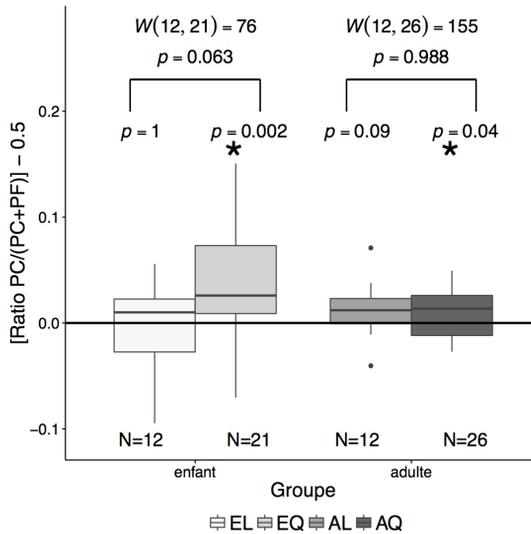


FIGURE 2: Graphique en boîte à moustache représentant le ratio des médianes des temps de réponse pour les essais des mots modifiés par la prononciation composée, dans les différentes conditions d'expérience (en laboratoire ou à l'extérieur) pour les deux groupes testés (adultes et enfants).

4 Discussion

Dès le plus jeune âge, la perception et la compréhension du langage est un enjeu important, car les enfants doivent rapidement apprendre une langue et interagir avec des personnes diverses présentes dans leur environnement. Nous avons ainsi étudié la compréhension des mots chez des enfants âgés de 24 à 48 mois, en parallèle avec des adultes.

Nous avons répliqué des études antérieures montrant que les prononciations non-standard ont un coût cognitif : les adultes et les enfants sont ralentis quand un mot n'est pas prononcé de façon standard, et ce même en présence de signes que l'interlocuteur peut ne pas être précis, puisqu'il a un accent étranger. Ce qui constitue la principale nouveauté de notre étude est que nous avons mesuré ce coût non seulement en laboratoire, mais aussi dans des environnements quotidiens pour nos participants. Nos résultats montrent que la perception de la parole accentuée n'est pas affectée différemment lorsqu'un étudiant est testé à la bibliothèque ou au laboratoire ; ni lorsqu'un enfant est testé à la crèche ou à l'école par rapport à une cabine insonorisée et isolée. Ceci pourrait indiquer que, bien que le contexte environnemental puisse biaiser la perception de façon globale (par ex. Hay & Drager, 2010), il ne change pas les mécanismes basiques de compréhension de mots.

Bien sur, il est possible qu'une différence existe à travers les environnements de test, mais elle serait trop petite pour être détectée avec la procédure employée ici. En effet, nous avons utilisé un plan expérimental où le facteur contexte varie à travers les groupes, ce qui résulte à moins de puissance statistique qu'un dessin intra-sujet. Bien que des recherches ultérieures pourraient corriger cette

limitation, nous attirons l'attention sur le fait que la tendance n'est pas vers plus de tolérance pour les prononciations inattendues dans un environnement quotidien, mais l'inverse.

Cette étude, comme d'autres, pourraient profiter de la validation d'une technique encore très peu utilisée en recherche fondamentale, le test sur tablette tactile, dont nous avons fait usage. Le fait que nous avons pu détecter le ralentissement en perception induit par une prononciation inattendue même dans un environnement peu contrôlé, suggère que cette technique pourrait nous permettre de rendre nos protocoles plus facilement transportables, et d'utiliser un support stable dans des lieux spécialement adaptés aux tests scientifiques (crèches, écoles, etc.)

Remerciements

Ce travail a été possible grâce au support économique et institutionnel de ANR-10- LABX-008 IEC (à travers de leur fonds d'aide aux chercheurs arrivants), ANR-10-IDEX-0001-02 PSL* (par ses Action Incitatives), et ANR-14-CE30-0003 MechELex.

Références

- BOERSMA, P., WEENINK, D. (2015). Praat, a system for doing phonetics by computer. Amsterdam.
- CRISTIA, A., SEIDL, A., VAUGHN, C., SCHMALE, R., BRADLOW, A., FLOCCIA, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in psychology*, 3, 479.
- FLOCCIA, C., DELL LUCHE, C., DURRANT, S., BUTLER, J., GOSLIN, J. (2012). Parents or community? Where do 20-month-olds exposed to two accents acquire their representations of words? *Cognition* 124, 95-100.
- GIRARD, F., FLOCCIA, C., GOSLIN, J. (2006). Familiarité aux accents régionaux et identification de mots. *Actes des JEP 2006*, 449-452.
- HAY, J., DRAGER, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865-892.
- JOHNSON, K., STRAND, E. A., D'IMPERIO, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359-384.
- MAYE, J., ASLIN, R. N., TANENHAUS, M. K. (2010). The Weckud Wetch of the West: Lexical adaptation to a novel accent. *Cognitive Science* 33, 543-562.
- R DEVELOPMENT CORE TEAM. (2008). R: A language and environment for statistical computing. R foundation for Statistical Computing. Vienna, Austria.
- SCHMALE, R., SEIDL, A., CRISTIA, A. (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental science*, 18(4), 664-670.
- VAN DER FEEST, S. V., JOHNSON, E. K. (2015). Input-driven differences in toddlers' perception of disappearing phonological contrast. *Language Acquisition*, 1-23.
- VAN HEUGTEN, M., JOHNSON, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, 143(1), 340.