

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/338854271>

Transfer of sensorimotor learning and speech units' development

Thesis · December 2019

CITATIONS

0

READS

8

1 author:



[Tiphaine Caudrelier](#)

GIPSA-lab

8 PUBLICATIONS 17 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Speech Unit(e)s [View project](#)

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : PCN - Sciences cognitives, Psychologie et
Neurocognition

Arrêté ministériel : 25 mai 2016

Présentée par

Tiphaine CAUDRELIER

Thèse dirigée par **Amélie ROCHET-CAPELLAN**, CNRS,
et co-dirigée par **Pascal PERRIER (EDISCE)**, Professeur, G-INP,
et par **Jean-Luc SCHWARTZ (EDISCE)**, CNRS

préparée au sein du **Laboratoire Grenoble Images Parole
Signal Automatique**
dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et
l'Environnement (EDISCE)**

Transfert d'apprentissage sensorimoteur et développement des unités de parole

Transfer of sensorimotor learning and speech units' development

Thèse soutenue publiquement le **22 Mai 2019**,
devant le jury composé de :

Madame Amélie ROCHET-CAPELLAN

Chargée de recherche, CNRS, Directrice de thèse

Monsieur Pascal PERRIER

Professeur, Grenoble-INP, Co-directeur de thèse

Monsieur Jean-Luc SCHWARTZ

Directeur de Recherche, CNRS, Co-directeur de thèse

Madame Marilyn VIHMAN

Professeure, Université de York, Royaume Uni, Rapporteure

Madame Chotiga PATTAMADILOK

Chargée de recherche, CNRS, Rapporteure

Madame Lucie MENARD

Professeure, Université du Québec à Montréal, Canada, Examinatrice

Madame Elsa SPINELLI

Professeure, Université Grenoble Alpes, Présidente du jury



Financement

Cette thèse a été entièrement financée par l'ERC Speech Unit(s) (FP7/2007-2013, Grant 339152) de Jean-Luc Schwartz, que je tiens à remercier pour avoir accepté de financer cette thèse.

Mon séjour de recherche à Montréal a été en partie financé par la Région Auvergne-Rhône-Alpes, grâce à la bourse Explora Doc'.

Abstract

Speech motor control has traditionally been studied apart from other cognitive processes underlying speech production, since first cognitive theories presented the brain as a set of relatively independent modules (Fodor & Pylyshyn, 1988), taken apart from the body. However developments in embodied cognition (Varela, Thompson, & Rosch, 1991), grounded cognition (Barsalou, 2008) and dynamic systems (L. B. Smith & Thelen, 2003) occurred in the last three decades underline that cognition cannot be considered separately from a body and its environment. These frameworks constitute an inspiration for this thesis and a motivation to study motor control and sensorimotor processes in relation to other cognitive processes. Whether linguistic structures are grounded in sensorimotor processes will be an underlying question.

A spoken message can be decomposed into sequences of linguistic units hierarchically structured. We argue that these speech units are grounded in sensorimotor representations, associating linguistic structures with auditory and motor information. Do these units correspond to words? Syllables? Phonemes? To probe the building blocks of speech production, we propose to use a paradigm of auditory-motor learning based on auditory feedback perturbation (Caudrelier & Rochet-Capellan, in press). This paradigm actually enables to change specific internal sensorimotor representations in speakers. Adaptation induces updating sensorimotor representations underlying the production of the training item. We assume that if this change affects the pronunciation of another word, it means that this word uses some of these updated representations. Thus, transfer patterns may reveal the structure of representations at stake.

A first study in adults shows that transfer of auditory-motor learning occurs at word, syllable, and phoneme levels in parallel (Caudrelier, Schwartz, Perrier, Gerber, & Rochet-Capellan, 2018). These observations suggest that all these units may co-contribute to the organization of speech articulation in adult speakers. Experimental results are discussed in the light of existing theories and models of speech production. A second experiment suggests that whether a speaker reads a word aloud or names a picture may have an influence on the transfer of auditory-motor learning (Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2018). A third study in 4- to 5-year-old and 7- to 8 year-old children investigates whether phoneme sensorimotor representations may emerge during reading acquisition, or prior to it (Caudrelier et al., in press). The observed transfer patterns suggest that phoneme representations emerge before reading acquisition, as a consequence of speech experience. Moreover, we found a relationship between adaptation to auditory perturbation and phonological awareness scores in both age groups. This suggests a link between sensorimotor representations and more explicit phonological representations. The potential causal or predictive nature of this link is discussed.

Overall, this work exploits an original and fruitful tool to probe speech representations and study their development. It may have clinical implications with regards to speech rehabilitation, as well as developmental dyslexia. It also highlights connections between speech sensorimotor level and higher linguistic and contextual levels that further question the nature of speech representations.

Keywords: speech production, phonological representations, sensorimotor learning, transfer of learning, auditory perturbation, phonological awareness

Résumé

Le contrôle moteur a traditionnellement été étudié séparément des autres processus cognitifs qui sous-tendent la parole, dans la lignée de théories de la cognition présentant le cerveau comme un ensemble de modules relativement indépendants (Fodor & Pylyshyn, 1988). Cependant les recherches autour de la cognition incarnée (Varela et al., 1991) et située (Barsalou, 2008), ainsi que des systèmes dynamiques (Smith & Thelen, 2003), menées ces trois dernières décennies soulignent que la cognition ne peut pas être considérée séparément d'un corps et de son environnement. Ce cadre constitue une source d'inspiration pour cette thèse et une motivation pour étudier les processus sensorimoteurs de la parole en lien avec les autres processus cognitifs.

La parole peut-être décomposée en séquence d'unités linguistiques structurées sous forme d'une hiérarchie. Nous soutenons que ces unités sont ancrées dans des représentations sensorimotrices, associant une structure linguistique avec des informations perceptives et motrices. Ces unités correspondent-elles à des mots ? Des syllabes ? Des phonèmes ? Pour sonder les représentations assurant l'articulation de la parole, nous proposons d'utiliser un paradigme d'apprentissage auditorimoteur basé sur la perturbation du retour auditif (Caudrelier & Rochet-Capellan, sous presse). Ce paradigme permet de modifier chez un locuteur des représentations sensorimotrices spécifiques, les représentations qui sous-tendent la production d'un item d'entraînement, par exemple un mot. Nous faisons ainsi l'hypothèse que si cette modification affecte la prononciation d'un autre mot, cela veut dire que la production de ce mot s'appuie sur une partie de ces représentations. Ainsi, l'observation du transfert d'apprentissage permet de révéler la structure de représentations qui assurent la production de parole.

Une première étude chez l'adulte montre que le transfert d'apprentissage auditorimoteur a lieu à la fois aux niveaux du phonème, de la syllabe et du mot (Caudrelier, Schwartz, et al., 2018). Ces observations suggèrent que ces unités co-contribuent à l'articulation de la parole chez l'adulte. Les résultats sont mis en perspective par rapport aux théories et modèles de production de parole. Une 2ème expérience suggère que la modalité de présentation du stimulus (un mot à lire ou une image à dénommer) peut influencer le transfert d'apprentissage auditorimoteur (Caudrelier, Perrier, et al., 2018). Une 3ème étude chez des enfants de 4-5 ans et de 7-8 ans montre que les représentations du phonème émergent avant l'acquisition de la lecture (Caudrelier & Rochet-Capellan, sous presse). De plus, un lien entre adaptation à la perturbation auditive et conscience phonologique est mis en évidence dans les deux groupes d'âge. Le potentiel caractère prédictif ou causal de ce lien est discuté.

En conclusion, cette thèse exploite un outil original et productif pour explorer les représentations de la parole et étudier leur développement. Ce travail pourrait avoir des implications cliniques, pour la rééducation de la parole, et pour la dyslexie développementale. Il met en évidence des liens entre les niveaux sensorimoteurs, linguistiques et contextuels qui questionnent la nature des représentations qui sous-tendent la parole.

Mots-clés : production de parole, représentations phonologiques, apprentissage sensorimoteur, transfert d'apprentissage, perturbation auditive, conscience phonologique

Remerciements

Tout d'abord, je remercie les membres du jury, en particulier Marilyn Vihman, Chotiga Pattamadilok, et Elsa Spinelli pour leur lecture attentive de ma thèse, leurs questions pertinentes et leurs retours constructifs, qui vont me permettre de développer de nouvelles perspectives sur mes recherches. Merci aussi pour votre bienveillance lors de ma soutenance de thèse.

Merci à Amélie, pour avoir éveillé ma curiosité scientifique sur les phénomènes sensorimoteurs, les modèles multi-traces, les systèmes dynamiques, et ta culture « psycho » en général. Merci pour ta franchise, ton écoute, et aussi pour tes capacités de remise en question. Vu les difficultés qu'on a eues pour la rédaction de mon mémoire de Master, nos relations doctorante-encadrante ne s'annonçaient pas comme un long fleuve tranquille, surtout pour les périodes de rédaction. Et pourtant, malgré une petite crise en Septembre 2016, je trouve qu'on a évolué ensemble et qu'on s'en est vraiment bien sorties! Je te remercie d'avoir su lâcher-prise et m'avoir accordé petit à petit ta confiance, en particulier pour la rédaction de l'article de Cognition et du présent manuscrit.

Merci à Pascal, pour toutes les fois où l'on n'était pas d'accord, et qui m'ont exercée à défendre mon point de vue scientifique. Merci d'avoir permis de réaliser mon souhait d'orienter ma thèse sur le développement de l'enfant, en lançant la collaboration avec Lucie. Merci d'avoir été là, la porte de ton bureau toujours entrouverte en fin de journée, pour mes questionnements du soir. Merci aussi d'avoir « oublié » de me dire que ce n'était pas une bonne idée de tomber enceinte pendant la thèse ;-)

Merci à Jean-Luc, pour ta vivacité d'esprit et ta clairvoyance, qui ont éclairé mes réflexions, et ont permis que nos longues discussions à 4 aboutissent à quelque chose de constructif. Merci aussi pour ta disponibilité et ta gentillesse. Vous avez formé une belle équipe d'encadrement ! Merci à vous 3 pour m'avoir soutenue, encouragée, et vous être montrés flexibles et compréhensifs vis-à-vis de ma situation.

Merci à Lucie, pour son accueil à Montréal, sa bienveillance et son soutien indéfectible pendant cette période qui a sans doute été la plus difficile de ma thèse : la collecte de données avec des enfants, dans les écoles, et les crèches... avec en plus à la maison un petit garçon de 5 mois, plutôt têtue, qui ne voulait pas boire de biberon, et n'acceptait que le sein de sa mère. Merci à toute l'équipe du laboratoire de phonétique de l'UQAM, à Montréal, et en particulier à Camille, Cristina, et Lambert, pour leur accueil chaleureux, leur enthousiasme, et leur aide précieuse lors

de la passation des expériences avec les enfants. Merci aussi aux écoles et aux crèches qui nous ont accueillis pour faire passer la manip.

Merci à mes collègues du labo. Merci à Christophe et Coriandre pour leur soutien technique sur les manip. Merci au service informatique, en particulier à Mikaël et Olivier, qui ont mainte fois « sauvé » mon ordinateur et son contenu. Merci à Silvain pour ton aide gigantesque sur les statistiques, ta disponibilité, ta réactivité, ta patience, pour les discussions et confidences, et pour ton humour décalé en toutes circonstances. Merci à mes co-bureaux, Thibaut, Clémence, Antje, et Laurence, pour l'ambiance conviviale dans le bureau pour ma première de thèse, ambiance qui m'a manquée sur la fin, après le départ des unes et les entrevues en pointillé avec les autres. Merci à Juan (Jean-François) pour les longues discussions scientifiques et perso, les craquages, et pour l'amitié qui s'est construite au fil de cette expérience au labo. Merci à tous les permanents du DPC, en particulier de PCMD, notamment Maeva, Thomas, Sonia, Anne, Marion, Pierre ; vous formez une équipe sympathique et bienveillante. Merci aux thésards et post-docs du labo, Alexandre, Diandra, Guillaume, Angélique, Francesca, Maël, Rémi, Andrew, Branco, Omar, Mélaine, Vincent, pour les discussions animées à la cafèt ! Merci à ceux parmi vous qui ont participé à ma manip : Xavier, Rémi, Paul, Maël, Elie. Merci à Alexandra, « ma » stagiaire, qui s'est démenée pour développer un bel outil de détection de formants. Merci à l'équipe de foot GIPSA 2, pour les moments sportifs que nous avons partagés.

Merci à tou.te.s les participant.e.s (au total plus de 200, si on compte les pilotes et les loupés...), ainsi qu'à leurs parents (pour les mineurs), qui ont bien voulu se prêter au jeu et passer une des manips les plus chiantes qui existent. Parmi eux, un merci à mes potes qui se sont déplacés pour y participer : Rachel, Karine, Cécilia, Léa, Nathalie, Anne-Lise, Johan, Julien, Guigui, Ilan, Loïc, Sandrine, et Anne.

Merci à mes ami.e.s et potes, qui m'ont accompagnée pendant ces années, notamment ceux qui ont partagé mon quotidien : les colocs de Montbonnot au début de la thèse, les potes de Sainte Luce sur la fin. Merci à Claire, Flo et Julie pour les bouffes entre filles. Merci aux potes qui se sont intéressés à ma thèse au point d'assister à ma soutenance : Flo, Michaël, Rémi, Eugé. Et merci à tous les autres, pour les bons moments que nous avons partagés, au cours des 3 ans et demi qui viennent de s'écouler.

Merci à ma famille. En particulier, merci à mes parents, qui m'ont permis d'avoir confiance en moi et ainsi de me sentir libre de changer de voie professionnelle. Merci aussi de vous être occupés de Justin un jour par semaine, en étant flexible sur les horaires! Merci à mon frère, qui

a été enthousiaste à l'idée de ma reconversion dans la recherche, et que j'encourage à poursuivre ses propres envies professionnelles. Merci à ma sœur pour s'être souvent intéressée au contenu de ma thèse, et pour avoir été un soutien constant au fil des années. Merci aux 3 enfants de ma sœur, Tristan, Ninon et Basile, qui ont été des cobayes enthousiastes pour ma manip sur les enfants.

Merci à Joseph, de m'avoir encouragée à franchir le pas de la reconversion professionnelle, apporté son soutien tout au long de celle-ci, et supportée dans les moments de doutes et de découragement de ma thèse (les erreurs dans le programme de ma manip qui m'ont fait perdre des sujets à Montréal, et les révisions de l'article du JSLHR qui n'en finissaient pas, entre autre). Merci de m'avoir accompagnée à Montréal et de t'être occupé au mieux de Justin pendant cette période, y compris en me l'apportant à l'UQAM le midi pour que je puisse continuer à l'allaiter. Merci d'avoir été présent pour Justin dans les moments où j'avais besoin de me consacrer pleinement à ma thèse, notamment le dernier mois de rédaction du manuscrit.

A Justin, merci d'être une source constante d'émerveillement, en particulier de par le développement incroyable de ton langage. La précision avec laquelle tu manies la parole du haut de tes 2 ans et demi, ainsi que la finesse et la sensibilité qu'elle laisse transparaître, m'impressionnent beaucoup ! Merci aussi d'avoir rythmé mes journées dès ma 2ème année de thèse... ça m'a obligée à viser l'efficacité !

Et merci à son petit frère, pour m'avoir fixé une date limite pour finir la rédaction de ce manuscrit, avec l'objectif de soutenir ma thèse avant qu'il ne sorte le bout de son nez... Tu m'as donné le cadre dont j'avais besoin pour clôturer tout ça, avec la perspective d'ouvrir une nouvelle page !

Table des matières

Financement	3
Abstract	5
Résumé	7
Remerciements	9
Table des matières	13
Table des illustrations.....	17
Avant-propos	19
Les phénomènes sensorimoteurs au cœur de la cognition.....	19
Les représentations sensorimotrices de la parole	20
Organisation et contributions de cette thèse	21
1. Les unités de parole.....	23
1.1 Introduction	23
1.2 Nature et niveau de granularité des unités en production de parole	24
1.2.1 Modèles psycholinguistiques hiérarchiques : phonème et/ou syllabe ?.....	24
1.2.2 Théories de l'exemplaire ou phonologie basée sur l'usage : phonème et/ou mot ?	28
1.2.3 Phonologie articulatoire et théorie de la sélection/coordination : syllabe et geste articulatoire.....	29
1.2.4 Modèles de contrôle moteur	31
1.2.5 Conclusion.....	33
1.3 Nature des unités en perception de la parole	34
1.3.1 La nature des unités dans les théories de la perception de la parole	34
1.3.2 Quel niveau de granularité des unités en perception ?.....	35
1.3.3 Conclusion.....	38
1.4. Développement des unités de parole chez l'enfant.....	39

1.4.1	En perception chez le bébé.....	40
1.4.2	En production : du babillage à l'âge adulte.....	42
1.4.3	Interactions production et perception.....	45
1.4.4	Développement phonologique et lexical.....	46
1.4.5	Conclusion.....	47
1.5.	Apprentissage de la lecture et développement de la conscience phonologique	48
1.5.1	La conscience phonologique	49
1.5.2	Interactions entre la lecture et le traitement de la parole.....	51
1.5.3	Dyslexie, représentations phonologiques et déficits sensorimoteurs	56
1.5.4	Conclusion.....	60
1.6.	Bilan des méthodes d'investigation des unités et projet de cette thèse	60
2.	Adaptation auditorimotrice : les perturbations de formants.....	63
2.1	Introduction	63
2.2	Vingt ans de recherche sur les perturbations de formants	65
3.	Etude 1 - Transfert d'apprentissage et unités de parole chez l'adulte.....	129
4.	Apprentissage sensorimoteur, transfert, et développement des unités phonologiques	145
4.1	Introduction	145
4.2	Etude 2 - Dénomination d'images ou lecture de mots : la modalité affecte-t-elle l'adaptation auditorimotrice et son transfert ?.....	147
4.3	Etude 3 - L'apprentissage sensorimoteur révèle des représentations phonémiques chez les enfants pré-lecteurs.....	154
4.4	Résultats complémentaires de l'étude 3 chez l'enfant.....	161
4.4.1	Conscience phonologique et adaptation	161
4.4.2	Ratio de transfert	164
4.5	Conclusion.....	169
5.	Discussion générale.....	171
5.1.	Rappel des résultats principaux	171

5.1.1 Un transfert d'apprentissage pour l'analyse des rôles du phonème, de la syllabe et du mot.....	172
5.1.2 L'influence de la modalité dans le transfert d'apprentissage	173
5.1.3 Les représentations phonologiques révélées par le transfert chez l'enfant	174
5.2. Relations entre adaptation auditorimotrice et conscience phonologique	175
5.2.1 Une relation de cause à effet ?.....	175
5.2.2 Niveau syllabique ou phonémique ?	177
5.3. L'effet « phonème »	178
5.3.1 Phonème, allophone, rime, trait phonétique et/ou geste articulatoire ?	178
5.3.2 L'émergence du phonème chez l'enfant pré-lecteur	180
5.4. Les effets lexicaux et contextuels.....	181
5.5. Comment ces différents niveaux d'unités s'organisent-ils ?	184
5.5.1 Coexistence des unités	184
5.5.2 Apprentissage global ou spécifique : flexibilité de l'organisation des unités ..	185
5.6. Conclusion générale	187
6. Annexe : Amplitudes de l'adaptation, du transfert et du post-effet en fonction de la durée de l'entraînement.....	189
7. Références	191

Table des illustrations

Figures

Figure 1-1: Représentation du réseau utilisé pour la production d'une phrase dans le modèle de Dell (1986) ..	24
Figure 1-2: Encodage phonologique dans le modèle de Levelt (extrait de Levelt & Wheeldon, 1994)	26
Figure 1-3: Schéma des thèmes abordés dans cette section consacrée au développement	40
Figure 1-4: Frise chronologique situant chaque sous-section de la partie développementale de ce chapitre 1 en fonction de l'âge, et de si elle est en relation avec la perception de la parole (en haut), la production de parole (en bas) ou les deux.	40
Figure 2-1: schéma de perturbations dynamiques (A) du geste du bras (Mattar & Ostry, 2007); (B) de la trajectoire de la mâchoire (Tremblay & al., 2003) et (C) du retour somatosensoriel (Ito & al., 2016). Ces figures sont extraites des articles correspondants.	64
Figure 2-2: Les formants. (A) Spectrogramme (représentation avec le temps en abscisse et la fréquence en ordonnée) d'une syllabe où les formants F1 et F2 sont surlignés. (B) Trapèze vocalique dans l'espace (F1,F2)	65
Figure 4-3: Schéma des relations entre parole, lecture et conscience phonologique, adapté du chapitre 1.....	146
Figure 4.4: Scores partiels aux tâches de conscience phonologique relatifs à (1) la syllabe et la rime et (2) le phonème, en fonction du groupe d'âge et de si les enfants sont adaptés ou non.....	162
Figure 4.5: Amplitude en début de phase (intercept du modèle) par phase et par groupe d'âge	166
Figure 5-1: Représentation schématique des effets phonème, syllabe et mot, en relation avec les amplitudes de transfert et de post-effet.....	173
Figure 5-2 : Modèle qualitatif représentant le cas où l'adaptation serait stable au cours du développement, et constituerait ainsi un prédicteur de la conscience phonologique.....	176
Figure 5-3: Modèle de dénomination de mots et d'images (extrait de Ferrand Grainger & Segui, 1994)	183
Figure 6-1: Amplitude de l'adaptation, du transfert et du post-effet, à l'intercept, selon la durée de l'entraînement avec perturbation auditive.	189

Tables

Table 4-1: Phonological awareness scores in % of correct responses by age group and by subtask (extraite de Caudrelier et al., sous presse, section 3.3)	162
Table 5-1: Synthèse des paramètres ayant varié dans les trois études présentées dans cette thèse.	171

Avant-propos

Les phénomènes sensorimoteurs au cœur de la cognition

Les sciences cognitives s'intéressent au fonctionnement de la pensée, avec une approche pluridisciplinaire mêlant neurosciences, psychologie cognitive, linguistique, philosophie, intelligence artificielle, ou encore anthropologie. Les sciences cognitives ont été dominées dans la 2^{ème} moitié du XX^{ème} siècle par le paradigme cognitiviste. Le cognitivisme considère que la pensée repose sur des calculs (ou opérations logiques) appliqués à des représentations symboliques abstraites (Pylyshyn, 1980). Selon ce paradigme, ces représentations sont *amodales*, c'est-à-dire indépendantes des modalités perceptives, telles que la vision ou l'audition. Les systèmes perceptifs sont chargés de recueillir les informations, qui sont ensuite traduites (ou plus précisément *transduites*) en représentations abstraites pour être traitées. Les systèmes perceptifs, dits « de bas niveau », sont programmés génétiquement, et opèrent de manière indépendante, selon une organisation modulaire (Fodor, 1985). Ils sont ainsi cantonnés à un rôle périphérique. Il en va de même pour le système moteur. Cette vision modulaire a été étendue à d'autres aspects de la cognition, comme le langage (Pinker, 1997). Ainsi, dans la logique du paradigme cognitivisme, les sciences cognitives ont longtemps laissé de côté les phénomènes sensorimoteurs, associés à la perception et à l'action, et qui donc transgressent les frontières des modules s'intéressant plutôt aux phénomènes dits de « haut-niveau », considérés comme au cœur de la cognition.

Ces trente dernières années, une vision radicalement opposée au cognitivisme a été développée, avec l'apparition de théories remettant le corps, et donc les systèmes sensoriels et moteurs (dits sensorimoteurs), au cœur de la cognition humaine. L'*enaction* (Varela et al., 1991), propose que la cognition émerge de l'interaction entre le corps et l'environnement, dans une forme d'auto-organisation. Elle découle des différentes capacités sensorielles et motrices du corps (cognition *incarnée*) ainsi que des caractéristiques de son environnement (cognition *située*) (Barsalou, 2008). Selon Wilson (2002), la cognition est au service de l'action. Le concept de cognition incarnée implique que les mécanismes de la pensée opèrent de façon modale, c'est-à-dire reliée aux modalités perceptives, et non sur la base de représentations abstraites. Ces mécanismes reposeraient notamment sur des simulations sensorimotrices, comme le suggère la découverte de neurones-miroirs chez le singe (Rizzolatti, Luppino, & Matelli, 1998; voir aussi

Rizzolatti, 2005). Ces neurones-miroirs s'activent à la fois quand un individu effectue une action, et quand cet individu perçoit cette action effectuée par un autre individu.

La théorie des systèmes dynamiques de Smith & Thelen (2003) place également les phénomènes sensorimoteurs et leur interaction avec l'environnement au cœur de la cognition, et de son développement. Elle repose notamment sur une série d'études sur le concept de permanence de l'objet chez le jeune enfant. Thelen, Schöner, & Scheier (2001) montrent que la permanence de l'objet n'est pas une connaissance acquise par l'enfant une bonne fois pour toutes, mais que les actions de l'enfant laissant entrevoir cette connaissance sont le fruit de processus sensorimoteurs en interaction et sous l'influence de caractéristiques spécifiques à la situation. Ces travaux suggèrent que les concepts, ou la connaissance, n'existent pas en dehors des processus dépendant du temps et du contexte qui produisent l'action (Thelen et al., 2001).

Les représentations sensorimotrices de la parole

Parmi les fonctions cognitives, le langage et la parole n'ont pas échappé à l'influence d'une vision cognitiviste et modulaire (Pinker, 1997). Une dichotomie historique distingue l'étude des phénomènes de haut niveau, supposés opérer avec des unités abstraites, comme la sémantique, et celle des phénomènes dits de bas niveau, ou sensorimoteurs, avec la phonétique, étudiant l'articulation et l'acoustique des sons de la parole, et le contrôle moteur. L'étude de la production de parole bénéficierait probablement d'une intégration de la psycholinguistique et de l'étude du contrôle moteur, qui a fait défaut jusqu'à récemment (Hickok, 2012).

En particulier, les modèles psycholinguistiques de production de parole proposent une vision formelle, avec la manipulation de représentations amodales (abstraites). Ces unités sont définies en termes de niveau de granularité linguistique (mots, syllabes, phonèmes). Les modèles psycholinguistiques s'arrêtent là où est supposée intervenir l'articulation de la parole, et ses représentations sensorimotrices. A l'inverse, la plupart des modèles de contrôle moteur de la parole s'intéressent aux phénomènes moteurs et sensoriels, en présupposant l'existence d'unités formelles, mais sans en questionner la nature, ni l'existence.

Dans cette thèse, nous nous placerons à l'interface entre la psycholinguistique et l'étude du contrôle moteur de la parole. Nous nous intéresserons aux phénomènes sensorimoteurs qui sous-tendent la parole et son développement. L'objectif de cette thèse est d'apporter un nouvel éclairage sur les débats concernant les unités de production de la parole, qui animent la recherche en psycholinguistique, grâce à l'utilisation d'une méthode expérimentale issue du

domaine du contrôle moteur : le paradigme d'adaptation à des perturbations en temps réel du retour auditif (modifications de la fréquence des pics d'énergie spectraux, appelés « formants »). L'originalité de cette méthode tient au fait qu'elle exploite les corrélats sensorimoteurs de l'unité, en perturbant les relations entre action et perception. Le paradigme d'adaptation permet ainsi de sonder spécifiquement les représentations assurant l'articulation de la parole, tandis que l'étude de la généralisation des stratégies mises en place pour contrecarrer les effets de la perturbation, ou transfert d'adaptation, permet de déterminer le niveau de granularité de ces représentations. Le raisonnement repose sur l'idée suivante : si un apprentissage réalisé pour une action donnée, par exemple la prononciation d'un mot donné, affecte - ou se transfère à - la réalisation d'une autre action, comme la production d'un autre mot, alors il existe des représentations communes impliquées dans ces deux actions. Ainsi, le transfert d'adaptation est notre outil d'exploration des représentations sensorimotrices qui sous-tendent la parole.

Organisation et contributions de cette thèse

Ce manuscrit s'ouvre par un chapitre d'état de l'art sur la nature des unités de parole et leur développement.

La 1^{ère} contribution majeure (chapitre 2), qui a fait l'objet d'un chapitre de livre, dresse un état de l'art du paradigme de perturbation de formants utilisé dans les études expérimentales de cette thèse (Caudrelier & Rochet-Capellan, sous presse). Cet état de l'art propose une synthèse de 77 articles de journaux ou actes de conférence. Après une synthèse des outils utilisés pour réaliser les perturbations de formants, une revue thématique détaillée explique les principaux résultats des études abordées.

La 2^{ème} contribution de cette thèse (chapitre 3) est une étude expérimentale qui explore chez l'adulte les unités de production de la parole, en contrastant pour la première fois, grâce au paradigme évoqué précédemment, 3 unités majeures proposées dans les grandes théories et modèles de production de parole: le phonème, la syllabe et le mot. Cette étude montre qu'il n'y a pas un seul niveau d'unité assurant l'interface avec l'articulation, mais plusieurs niveaux opérant conjointement (Caudrelier, Schwartz, et al., 2018).

La 3^{ème} contribution (chapitre 4) associe deux études expérimentales autour des relations entre unités de parole et lecture. L'une d'elle teste l'influence de la modalité de la présentation du stimulus à prononcer (lire un mot ou nommer une image) sur les phénomènes d'adaptation et

de transfert au cœur de cette thèse (Caudrelier, Perrier, et al., 2018). L'autre est une étude expérimentale explorant chez l'enfant l'émergence des unités de production de parole (Caudrelier et al., 2019). Elle vise en particulier à mieux comprendre les relations entre émergence du phonème en production de parole et apprentissage de la lecture. Elle montre de manière inédite que le phonème émerge chez l'enfant avant l'apprentissage de la lecture, probablement comme une conséquence de l'expérience du langage oral. De plus, elle établit un lien entre les représentations sensorimotrices de la parole, et une compétence étroitement liée à l'apprentissage de la lecture, la conscience phonologique.

Ces résultats sont discutés au regard des théories et modèles de production de parole, et du développement phonologique de l'enfant. De nouvelles perspectives de recherche basées sur le paradigme de transfert d'apprentissage sensorimoteur sont proposées, en particulier pour explorer les représentations sensorimotrices de populations spécifiques, comme les personnes dyslexiques, ou illettrés.

1. Les unités de parole

1.1 Introduction

La linguistique formalise les langues en identifiant des unités qu'elle organise de manière hiérarchique, l'unité la plus petite (mais aussi la plus générale) étant enchâssée dans une unité plus grande (mais aussi plus spécifique), et ainsi de suite : phonèmes, syllabes, mots, phrase. La combinaison de ces unités confère au langage humain sa propriété combinatoire. Parmi ces unités, l'unité phonologique est une catégorie abstraite, qui renvoie à une réalité physique (notamment acoustique) multiple. Ainsi, à un même phonème correspondent des propriétés phonétiques multiples.

La psycholinguistique associe des représentations mentales à ces unités. Ces représentations seraient elles-mêmes organisées de manière hiérarchique, faisant miroir à la hiérarchie des unités décrite en linguistique (Dell, 1986; Goldinger & Azuma, 2003). On emploiera tour à tour le terme d'unité de production de la parole, et de représentation phonologique. Dans la suite du manuscrit, l'unité de production de parole désigne une représentation à l'interface avec l'articulation de la parole, l'unité de planification motrice, comme expliqué plus précisément en section 1. Le terme de représentation phonologique, plus couramment utilisé dans les parties sur le développement, désigne plus généralement les représentations situées entre les niveaux sémantique et moteur.

Dans ce chapitre, nous nous intéresserons à la nature de ces représentations phonologiques, avec un accent particulier sur celles qui sont utilisées pour la production de parole. Dans la première partie, nous examinerons cette question au travers de modèles et théories de production de la parole, principalement chez l'adulte. Une deuxième partie s'intéressera aux relations entre production et perception de la parole, ainsi qu'à la littérature sur les unités en perception. La troisième partie est consacrée au développement des unités phonologiques chez l'enfant, en perception et en production. Dans une quatrième partie, nous nous pencherons sur l'influence de l'apprentissage de la lecture sur la parole, et le développement des unités phonologiques. En dernier lieu, nous reviendrons sur les différentes méthodes évoquées au cours de ce premier chapitre, permettant d'explorer la nature des unités de parole, et introduirons la méthode empirique que nous avons utilisée pour interroger l'unité de production de la parole. La littérature concernant ce paradigme sera détaillée dans le 2^{ème} chapitre de cette thèse.

1.2 Nature et niveau de granularité des unités en production de parole

Plusieurs champs disciplinaires ont vu émerger des théories de la production de parole. Dans cette section nous examinerons les propositions de ces théories concernant la question suivante : quelle est l'unité de base de la production de parole ? Et plus précisément, quel est son niveau de granularité ?

1.2.1 Modèles psycholinguistiques hiérarchiques : phonème et/ou syllabe ?

Comprendre et modéliser la production de parole est l'un des principaux objectifs de la psycholinguistique. Les premiers modèles dans ce domaine ont été ceux de Fry (1969) et Garrett (1975) dont les modèles ultérieurs se sont inspirés. Dell (1986) et Levelt (1989) ont chacun proposé une théorie, basée sur des données expérimentales et sur des corpus de parole. Elles reposent en particulier sur l'analyse des erreurs de production, les effets de fréquence sur le temps de réaction dans des tâches de production, ou encore sur des expériences de *priming*.

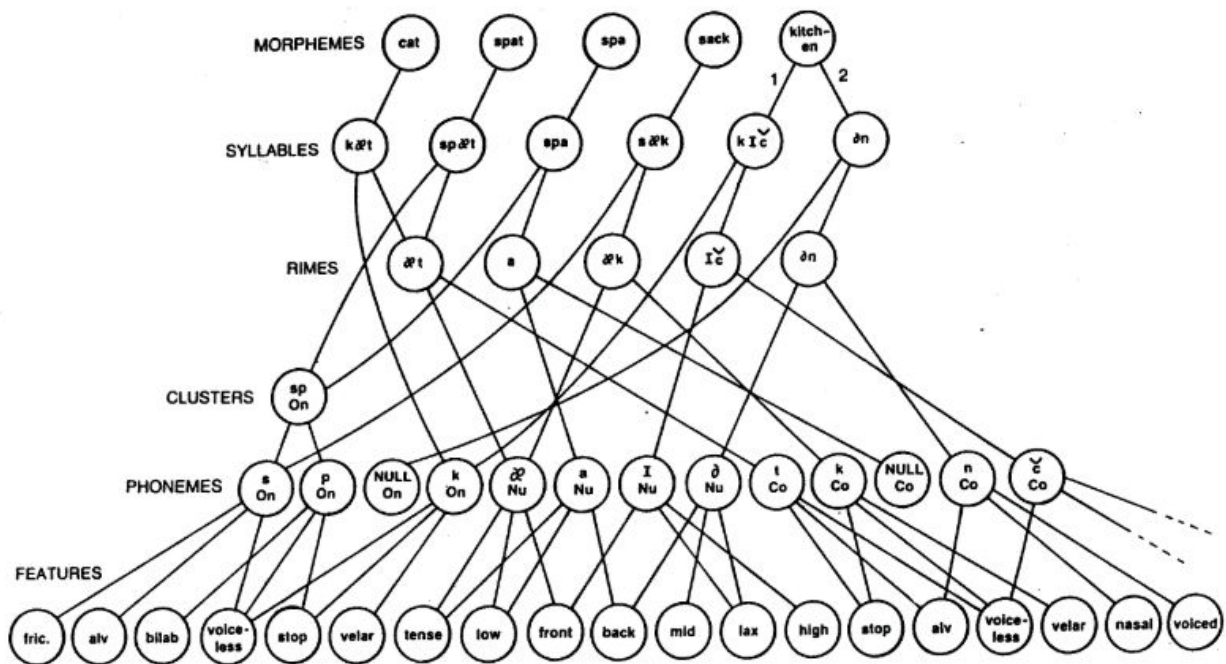


Figure 1-1: Représentation du réseau utilisé pour la production d'une phrase dans le modèle de Dell (1986)

Dell (1986) propose une théorie de la propagation de l'activation (*spreading activation theory*) pour la production de phrases, illustrée par la Figure 1-1. Elle s'intéresse à la fois aux processus d'encodage syntactique, morphologique et phonologique. L'encodage phonologique est le

processus de décomposition des unités sémantiques (mots ou morphèmes) en unités plus petites, qui pourront être traduites en commandes motrices, afin de prononcer le mot ou le discours souhaité. Le modèle concernant l'encodage phonologique est constitué de couches correspondant à un niveau dans une organisation hiérarchique d'unités linguistiques : morphèmes, syllabes, rimes/clusters, phonèmes, et même traits phonétiques. C'est l'activation des nœuds du réseau qui va permettre de sélectionner les unités à produire, dans l'ordre souhaité. L'activation se propage du haut vers le bas, du morphème vers le phonème, mais aussi dans le sens inverse, ce qui permet d'ailleurs d'expliquer certaines erreurs de production. Un système de file d'attente permet à chaque nœud d'être activé au bon moment. A chaque niveau, syntactique, morphologique ou phonologique, il existe des règles génératives qui fournissent un cadre, dans lequel les processus d'activation s'opèrent. Au niveau phonologique, les règles sont phonotactiques. Le cadre qui en découle est représenté par la syllabe, et ses différents constituants : l'attaque (*onset*), le noyau (*nucleus*), et la *coda*. Le début et la fin sont composés d'une consonne, d'un cluster de consonnes ou de l'élément nul, tandis que le noyau est généralement constitué d'une voyelle ou d'une diphtongue. Ainsi dans le modèle de Dell, la syllabe est le cadre (*frame*) et le phonème est le constituant (*content*). De ce point de vue, la théorie de Dell peut rappeler la théorie du Cadre/Contenu de MacNeilage (1998), que l'on développera plus tard dans ce chapitre.

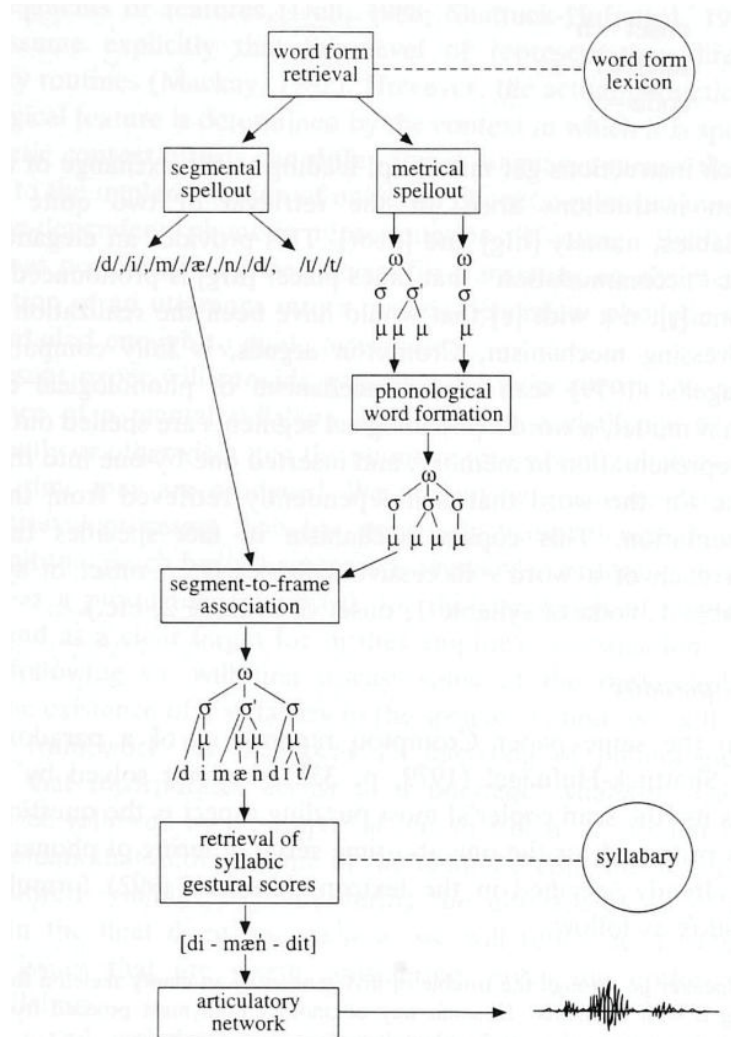


Figure 1-2: Encodage phonologique dans le modèle de Levelt (extrait de Levelt & Wheeldon, 1994)

Un des objectifs du modèle de Dell est de rendre compte de la variété et de la spécificité des erreurs de production. Malgré le fait que de nombreuses erreurs de production (ou *lapses*) sont des échanges ou substitutions de phonèmes (Shattuck-Hufnagel & Klatt, 1979), Dell affirme que le phonème n'est pas la seule unité de production. En effet, il remarque le rôle ambigu joué d'une part par la syllabe, et d'autre part par les traits phonétiques. Il y a très peu d'échanges de syllabes ou de traits phonétiques dans les corpus d'erreurs de production existants (Shattuck-Hufnagel & Klatt, 1979). Pourtant, la syllabe et les traits phonétiques jouent un rôle essentiel pour déterminer quel phonème peut être remplacé par quel autre phonème. Par exemple, les

phonèmes échangés partagent souvent des traits phonétiques, et ont la même place dans la structure syllabique.

Levelt propose un modèle de production de parole un peu différent, illustré par la Figure 1-2 (Levelt & Wheeldon, 1994; Levelt, 1989). Ce modèle part de l'intention de communication, pour aller jusqu'à l'articulation de la parole. Il inclut notamment une étape d'encodage phonologique qui produit un plan phonétique ou parole interne, qui précède l'articulation dans le conduit vocal. Comme Dell, Levelt fait une distinction entre le cadre, qu'il appelle *squelette*, et le contenu segmental. Contrairement à Dell, le cadre n'est pas la syllabe mais plutôt le lemme, qui correspond à l'architecture du mot entier. Lors de l'encodage phonologique, le formulateur décompose le mot en segments (ou phonèmes). Il y a ensuite une recombinaison en syllabes, qui constituent l'unité de base de l'articulation. Le fait que la syllabe intervienne en dernier, après la décomposition du mot en phonèmes, permet le phénomène de resyllabification que l'on observe par exemple en français, où une syllabe peut parfois être à cheval sur deux mots (Spinelli, McQueen, & Cutler, 2003). Les commandes motrices associées à chaque syllabe courante sont stockées en mémoire dans un syllabaire mental.

Le concept du syllabaire mental repose essentiellement sur des expériences mesurant le temps de réaction pour prononcer un mot, en fonction notamment du nombre de syllabes dont il est constitué et de leurs fréquences (Cholin, Levelt, & Schiller, 2006). Ce concept est aussi en cohérence avec l'observation selon laquelle les phonèmes correspondent à des gestes qui se recoupent et se recouvrent au sein de la syllabe. Des expériences d'amorçage, ou *priming* en anglais, ont également soutenu la syllabe comme unité de production de la parole (Ferrand, Segui, & Grainger, 1996). Le paradigme de priming consiste à faire apparaître un stimulus (appelé amorce, ou *prime* en anglais), puis à mesurer le temps de réaction pour prononcer une cible, en anglais *target*, qui à son tour est présentée juste après. Des variantes peuvent proposer d'autres tâches, comme une tâche de décision lexicale sur le stimulus cible. Une réduction du temps de réaction par la présence de l'amorce permet de déduire que l'amorce et la cible ont une représentation commune. Il a été montré que le temps de réaction est plus faible lorsque l'amorce et la cible ont la première syllabe en commun, que lorsqu'elles ont en commun une lettre en plus (ou en moins) que la première syllabe (Ferrand et al., 1996; mais voir aussi Perret, Bonin, & Méot, 2006).

En résumé, bien que Dell et Levelt proposent chacun une vision différente de l'encodage phonologique, leurs modèles ont en commun l'utilisation de hiérarchies d'unités, où la syllabe

et le phonème ont une place essentielle. L'information à chaque étape du modèle, notamment pour le passage du niveau phonologique à l'articulation, est caractérisée selon un niveau précis dans la hiérarchie des unités linguistiques. Néanmoins, Dell et Levelt sont sensibles aux limites de cette vision, et y apportent des nuances. La théorie proposée par Dell reste malgré tout assez floue sur les liens effectifs avec l'articulation. Chez Levelt, les syllabes fréquentes sont associées à des routines articulatoires utilisées pour la prononciation, mais il est aussi possible de reconstituer une syllabe moins fréquente à partir de ses constituants phonémiques.

1.2.2 Théories de l'exemplaire ou phonologie basée sur l'usage : phonème et/ou mot ?

Les théories de l'exemplaire offrent une perspective très différente de celle exposée dans la section précédente pour questionner la nature des unités de parole. La notion d'exemplaire a d'abord été développée en psychologie pour expliquer la catégorisation en général (Medin & Schaffer, 1978). Dans le domaine de la parole, ces théories ont été développées essentiellement en perception, et plus marginalement en production. Ces théories ne raisonnent pas en termes d'unités abstraites (ou de représentations) qui correspondraient par exemple aux nœuds du réseau proposé par Dell. Au lieu de catégories abstraites, ce sont toutes les expériences sensorielles et motrices, nommées traces, qui seraient stockées en mémoire. Ainsi, le lexique contiendrait des traces détaillées, incluant par exemple des détails acoustiques propres à la voix et la prononciation d'un locuteur donné (Bybee, 2002; Goldinger, 1996; Hawkins, 2003; Johnson, 1997; Klatt, 1979; Pierrehumbert, 2002). Cette vision repose notamment sur la mise en évidence d'effets de détails de surface des stimuli expérimentaux (ou en anglais *surface details*, dans Goldinger, 1996) aussi désignés parfois par « effet de correspondance perceptive » (en anglais *perceptual match effect*, dans Reder, Donavos, & Erickson, 2002).

Néanmoins, plusieurs points de vue diffèrent au sein des théories de la parole basées sur l'exemplaire. Pierrehumbert (2001) fait l'hypothèse que les différents exemplaires de segments (phonèmes ou allophones) sont stockés ensemble, et que la distribution de probabilité qui en découle forme une catégorie. A l'inverse, Välimaa-Blum (2009) soutient que le phonème ne peut être stocké indépendamment du mot, ni aucune catégorie sous-lexicale (dénuée de sens). En effet, si deux mots diffèrent seulement par un phonème, la reconnaissance de ce phonème dans l'un de ces deux mots est influencée par le contexte dans lequel est placé le mot, son sens. Ainsi le phonème est relié au sens. Or selon Välimaa-Blum il semble inutile d'associer un sens au phonème. De plus, la prononciation des phonèmes dépend aussi énormément du contexte dans lequel il est placé, pas seulement le contexte phonétique, qui peut avoir une influence pour

des raisons mécaniques évidentes, mais aussi le contexte socio-linguistique et la fréquence des mots par exemple. Ces observations soutiennent l'idée que l'on garde en mémoire les expériences passées liées aux mots ou groupes de mots (Välímää-Blum, 2009).

Dans ce cadre, la production est reliée à la perception. Le modèle Trace (McClelland & Elman, 1986) de la perception de la parole repose sur l'idée que l'on garde une trace de l'expérience sensorielle associée à chaque énoncé perçu. Quand un énoncé est prononcé ou entendu, il réactive les traces gardées en mémoire qui ont un élément commun avec lui. Ce modèle s'inspire de modèles multitraces de la mémoire (Hintzman, 1986) qui considèrent que tout événement perçu laisse une trace sensorielle, et chaque événement nouveau vient rentrer en résonance avec les traces existantes. C'est ce processus associé à la combinaison de traces qui permet à la catégorisation (ou la conceptualisation) d'émerger. La dimension sensorimotrice de cette conception de la mémoire basée sur des traces d'expérience a été développée plus récemment (Versace, Labeye, Badard, & Rose, 2009).

1.2.3 Phonologie articulatoire et théorie de la sélection/coordination : syllabe et geste articulatoire

La phonologie articulatoire (Browman & Goldstein, 1992) affirme que l'unité de base de la production de parole est le geste articulatoire, défini comme une action discrète de constriction glottique (en lien avec le voisement) ou supraglottique (en lien avec le lieu d'articulation dans le conduit vocal). Le geste correspond à un événement spécifié de manière spatiotemporelle, et les unités phonologiques (dont les phonèmes) seraient constituées d'un assemblage de gestes articulatoires – les atomes – superposés et organisés entre eux, comme une molécule.

Ces gestes articulatoires se manifesteraient chez l'enfant d'abord lors de la période du babillage. Cette période se caractérise en effet par une exploration motrice. Puis, lors des premiers mots, l'enfant assemblerait ces gestes (de manière souvent imprécise) pour produire un mot, ou signifiant, associé à un signifié. Progressivement, une meilleure synchronisation entre les gestes est acquise, permettant la constitution progressive d'unités phonologiques, en particulier la syllabe.

La syllabe est vue comme un motif spécifique de coordination des gestes (Browman & Goldstein, 1995b), qui s'accompagne de la notion de c-center. Le c-center est défini comme le point temporel central de la consonne initiale d'une syllabe, ou du cluster de consonnes initial (Browman & Goldstein, 1988). Ce c-center est coordonné de manière précise avec le geste du noyau (ils sont en phase). Par contre la coordination du noyau avec la coda est moins stricte, et

est le plus souvent en opposition de phase. Cette vision de la syllabe a l'avantage d'expliquer certaines variations dans la prononciation d'un phonème (différence de prononciation d'une consonne suivant si elle est placée en début, *onset*, ou en fin de syllabe, *coda*) sans faire appel à une structure hiérarchique prédéfinie (comme c'est le cas dans le modèle de Dell par exemple). Il s'agit d'une structure auto-organisatrice, née de l'interaction entre les contraintes cognitives et physiques (Browman & Goldstein, 1995a). Ainsi, l'unité phonologique émergerait au cours du développement à partir des propriétés physiques du système.

Plus généralement, cette vision s'oppose à la conception générativiste que la production de parole correspond à la mise en séquence d'unités phonologiques abstraites, non spécifiées phonétiquement, qui sont ensuite réalisées de manière approximative et variable suivant le contexte. Le geste articulatoire est au contraire, ici, à la fois l'unité cognitive manipulée mentalement et l'unité physique du mouvement dans le conduit vocal.

Dans le prolongement de la phonologie articulatoire, Tilsen (2016) considère également que l'unité de base de la production de parole est le geste articulatoire, et développe une théorie expliquant comment se construisent les assemblage de gestes articulatoires, ou unités phonologiques. Selon la théorie de la sélection/coordination, cette construction est effectuée grâce à deux mécanismes complémentaires, qui engendrent deux types de contrôle des articulateurs : la compétition et la coordination. La compétition permet la sélection des mouvements à produire dans le bon ordre, et la coordination assure le contrôle temporel précis de l'exécution des gestes (Tilsen, 2013, 2016). C'est un des points qui différencient cette vision de celle de la phonologie articulatoire, qui se base uniquement sur la coordination. Dans le développement de la parole, le mécanisme de sélection précéderait celui de coordination, qui dépend de l'utilisation du feedback sensoriel, grâce à des modèles internes (cette notion sera expliquée dans la sous-section suivante). Cette proposition se base notamment sur l'étude de la durée segmentale et la coarticulation des syllabes CV et VC (Consonne-Voyelle et Voyelle-Consonne) au cours du développement de l'enfant (Tilsen, 2016).

La phonologie articulatoire, et, dans son prolongement, la théorie de la sélection/coordination, proposent des mécanismes physiques qui permettraient l'émergence des unités de parole au cours du développement, contrairement aux modèles cités plus haut qui postulent une hiérarchie d'unités statique (Levelt, 1989 et Dell, 1986).

1.2.4 Modèles de contrôle moteur

Les modèles de contrôle moteur de la parole se situent au niveau du lien entre la phonologie et l'articulateur tel que défini par Levelt. Ces modèles ne cherchent pas à tester des hypothèses sur les unités linguistiques qui servent d'interface entre l'encodage phonologique et l'articulation elle-même. Ils se contentent en général d'en postuler l'existence. Néanmoins, ils permettent d'apporter un éclairage sur la nature même de ces unités, aux niveaux perceptif et moteur. Une revue détaillée des modèles de production de la parole est disponible dans la thèse de Patri (2018). Le lecteur trouvera également de plus amples informations sur la notion de modèle interne dans la thèse de Barbier (2016).

Le modèle Directions Into Velocities of Articulators (DIVA) décrit les interactions sensorimotrices intervenant dans l'articulation de la parole (Tourville & Guenther, 2011). L'un des éléments-clé de ce modèle est la carte des sons de parole (*Speech Sound Map, SSM*). Ces sons de parole sont mis en correspondance avec l'enchaînement de commandes motrices qui permettent de les réaliser, précisées sous la forme de vitesses des articulateurs. La mise en relation des sons de parole avec l'articulation correspond à ce que Levelt désigne par le terme d'encodage phonétique (voir Levelt & Wheeldon, 1994; Levelt, Roelofs, & Meyer, 1999). Cette carte des sons assurant la correspondance entre phonologie et articulation pourrait contenir aussi bien les syllabes fréquentes, comme proposé par Levelt, que des unités infra-syllabiques (phonèmes) ou multisyllabiques (mots, phrases) (Tourville & Guenther, 2011). Cependant, en pratique le modèle computationnel se base sur une cadre syllabique (ex : type CVC) associé à un contenu phonémique (Bohland, Bullock, & Guenther, 2010). Comme chez Levelt, les syllabes fréquentes sont directement représentées dans le SSM, et les syllabes moins fréquentes peuvent être recomposées à partir des constituants phonémiques et de leur place dans la structure syllabique (Bohland et al., 2010).

Le modèle GEPPETO (pour "GEstures shaped by the Physics and by a PErceptually Oriented Targets Optimization") associe un but moteur à chaque phonème de la séquence de parole à produire (Patri, Diard, & Perrier, 2015; Perrier, Ma, & Payan, 2005). Ces buts moteurs sont définis par une région de l'espace acoustique. L'articulation est alors vue comme une tâche d'interpolation de ces buts moteurs mis en séquence. En ce sens, les unités de GEPPETO sont des unités de planification car elles ne définissent pas de commandes motrices à elles-seules : c'est la séquence d'unités de planification qui permet de définir les commandes motrices, et la

trajectoire des articulateurs résulte de l'interaction entre les commandes motrices associées aux buts moteurs et les caractéristiques biomécaniques des articulateurs.

Le modèle de production de parole State Feedback Control, dit SFC (Houde & Nagarajan, 2011), propose que le contrôle de la parole se base sur un mécanisme permettant d'estimer l'état dynamique actuel du système (i.e. du conduit vocal). SFC ne fait pas d'hypothèse particulière au sujet des unités. Les corrélats neuronaux discutés sont valables essentiellement pour la production de phones ou de syllabes, mais le modèle serait extensible à d'autres unités plus larges.

Hickok (2012) propose une extension du modèle SFC, nommée HSFC (Hierarchical State Feedback Control). Dans ce modèle, il souhaite incorporer la vision psycholinguistique, avec son encodage phonologique sous forme de hiérarchie d'unités linguistiques, et la vision issue du contrôle moteur, inspirée du SFC, plus focalisée sur les liens sensorimoteurs. Dans ce modèle, les syllabes, ou plus précisément des buts moteurs qualifiés de « haut niveau », sont contrôlées au niveau acoustique, tandis que les phonèmes, ou cibles de plus « bas niveau », sont contrôlés au niveau somatosensoriel. Une nuance est toutefois apportée concernant cette dichotomie. Contrairement aux consonnes occlusives (par exemple, /b/ ou /d/) qui ont besoin d'être intégrées à une syllabe pour engendrer un son, les voyelles ainsi que certains types de consonnes, telles que les fricatives (/f/), ont une conséquence acoustique à elles-seules. Ainsi, ces phonèmes seraient définis à la fois sur les plans acoustique et somatosensoriel. Le modèle ne propose donc pas une hiérarchie purement basée sur les unités linguistiques, mais plutôt sur ce qu'il désigne comme des unités de contrôle : les points extrêmes des ouvertures ou fermetures du conduit vocal étant sous contrôle somatosensoriel, tandis que les trajectoires entre ces points seraient définies acoustiquement (Hickok, 2012).

Au-delà des unités linguistiques utilisées, un des points fondamentaux des modèles de contrôle moteur est la manière d'incorporer la perception dans le contrôle de la production de parole. La plupart de ces modèles repose sur deux mécanismes de contrôle, sous-tendus par ce qu'on appelle des *modèles internes* (Wolpert, Miall, & Kawato, 1998). Le *feedforward controller* est basé sur le modèle dit *inverse*, qui associe à un but sensoriel (acoustique et/ou somatosensoriel) un ensemble de commandes motrices permettant l'atteinte de ce but. Ce type de contrôle permet la réalisation de séquences motrices rapides sans l'aide du retour sensoriel, plus lent. Il ne permet néanmoins pas de correction en cas d'erreur. L'autre type de contrôle est le *feedback controller*, qui utilise l'erreur entre le retour sensoriel reçu (i.e. auditif ou somatosensoriel) et

ce qui était prévu pour ajuster les commandes motrices. Ces deux types de contrôle sont adoptés dans la plupart des modèles de production de parole, et de contrôle moteur en général (voir les thèses de Patri, 2018, et Barbier, 2016).

Ainsi les modèles de production mettent en avant le rôle fondamental de la perception dans la spécification des buts de la production de la parole, avec l'utilisation du retour sensoriel, en particulier le retour auditif. Ces modèles suggèrent que les unités de base de la production sont de nature sensorimotrice. Réciproquement, on peut se demander dans quelle mesure la production intervient dans la perception.

1.2.5 Conclusion

Dans les différentes théories évoquées, issues de divers champs disciplinaires, l'unité de base de la production de parole est tour à tour le phonème, la syllabe, le geste articulatoire, ou le mot. On remarque que les modèles psycholinguistiques sont basés sur une hiérarchie d'unités, qui interviennent successivement dans le processus de production de parole. Cette hiérarchie est fixe et ses mécanismes de construction ne sont pas précisés. D'autres théories suggèrent de possibles mécanismes de construction des unités au fil de l'expérience de la parole : dans la théorie de l'exemplaire c'est l'accumulation de traces mnésiques qui constitue l'unité ; dans la phonologie articulatoire et la théorie de la sélection/coordination, les unités émergent entre autres de propriétés physiques du corps (McClelland & Elman, 1986; Tilsen, 2016). Nous nous intéresserons plus particulièrement au développement des unités dans les sections 3 et 4 de ce chapitre.

A l'évocation de ces théories, on entrevoit aussi un autre débat concernant la nature des unités de production de parole. Bien que centrées que sur la production, certaines théories sont en étroite relation avec la perception. C'est le cas des théories de l'exemplaire (ou phonologie basée sur l'usage), ou dans une moindre mesure la phonologie articulatoire, qui a été développée dans le même contexte scientifique que la théorie motrice de la perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Ainsi, ces unités sont-elles spécifiques à la production de parole, ou sont-elles en relation avec les unités de la perception ? Quelles sont justement les unités considérées en perception de la parole ?

1.3 Nature des unités en perception de la parole

Dans la section précédente, nous avons décrit les unités de production à travers différents modèles et théories de la production de parole, en partant des structures linguistiques pour aller vers le fonctionnement des unités et les liens perceptuo-moteurs qu'elles instancient. Dans cette section, nous procéderons à l'évolution inverse pour explorer la nature des unités en perception de la parole. Nous nous intéresserons d'abord aux théories de la perception pour éclairer la nature des unités à travers les relations perception-production. Puis nous nous pencherons sur les différents paradigmes qui ont été utilisés pour répondre à la question : quelles unités linguistiques interviennent dans la perception de la parole ?

1.3.1 La nature des unités dans les théories de la perception de la parole

Nous nous intéressons dans un premier temps aux théories de la perception de la parole, dans la mesure où elles peuvent nous éclairer sur les relations perception – production. Pour de plus amples informations sur ces théories, le lecteur pourra se référer à la thèse de Barnaud (2018). La question des unités en perception est posée sous la forme : quelle est la nature de l'invariant de la parole, c'est-à-dire l'indice qui possède une relation stable (potentiellement bijective) avec le phonème correspondant ?

La théorie motrice (Liberman et al., 1967) part du prémisses qu'il est très difficile d'extraire un invariant acoustique. En particulier, le signal acoustique d'une consonne occlusive (e.g. /d/) dépend de la voyelle qui suit. Selon cette théorie, les unités (ou invariants) sont les gestes intentionnels, c'est-à-dire les commandes motrices destinées à l'articulation de la parole (Liberman & Mattingly, 1985). Dans cette vision, les unités de la perception et de la production de parole sont les mêmes. La théorie réaliste directe considère elle aussi un lien étroit entre perception et production, mais propose que ce soient les configurations articulatoires effectivement atteintes (et non intentionnelles) qui sont détectées lors de la perception de la parole (Fowler, 1986).

A l'inverse, les points de vue auditifs sur la perception (Blumstein & Stevens, 1979; Diehl, Lotto, & Holt, 2004; Stevens & Blumstein, 1978) tentent de montrer qu'il existe un invariant acoustique, même pour les consonnes occlusives, et en concluent que l'enfant comme l'adulte se basent uniquement sur le signal acoustique pour en extraire les unités de parole, les phonèmes, sans devoir faire appel à des connaissances articulatoires ou motrices sous-jacentes.

Enfin, les théories sensorimotrices proposent que la perception de la parole se base à la fois sur des informations acoustiques et motrices (Schwartz, Basirat, Ménard, & Sato, 2012; Schwartz, Berthommier, & Savariaux, 2004; Skipper, Devlin, & Lametti, 2017; Skipper, Van Wassenhove, Nusbaum, & Small, 2007). Ainsi grâce à l'imitation, l'enfant établirait des cartes sensorimotrices, reliant les représentations sensorielles et motrices de la parole. Cette notion de carte est d'ailleurs présente dans la plupart des modèles de production évoqués ci-dessus. Pour ces visions perceptuo-motrices, les unités en perception et en production sont donc reliées, sans être pour autant nécessairement confondues. Le modèle COSMO propose une définition de ces unités et de leurs liens en termes de distribution de probabilité (Barnaud, Bessière, Diard, & Schwartz, 2018; Moulin-Frier et al., 2015).

1.3.2 Quel niveau de granularité des unités en perception ?

En parallèle de ce débat sur la nature sensorielle vs. articulatoire/motrice des unités de parole, qui considère souvent le phonème comme unité de base de la perception, d'autres recherches ont questionné l'unité de perception de la parole au sens du niveau dans la hiérarchie des unités linguistiques (Goldinger & Azuma, 2003). Comme la perception n'est pas au centre de cette thèse, nous ne décrivons que brièvement la littérature sur ce sujet, et nous ne rentrerons pas dans le détail des principaux modèles (Luce, Goldinger, Auer, & Vitevitch, 2000; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000). Nous nous attacherons plutôt à passer en revue les principales unités linguistiques considérées, et les études expérimentales qui les soutiennent. En effet, les paradigmes expérimentaux utilisés en perception pourraient se révéler source d'inspiration pour l'étude des unités en production. Nous détaillerons en particulier les paradigmes d'adaptation sélective et d'apprentissage perceptif, qui représentent un équivalent en perception du paradigme que nous allons utiliser pour étudier les unités en production.

Si le phonème a été l'unité historiquement la plus considérée, la syllabe a connu un regain de popularité, et ces deux unités se sont longtemps disputé le devant de la scène. L'observation de Liberman et coll. évoquée précédemment, selon laquelle il n'existe pas d'invariant acoustique pour les consonnes occlusives, peut donner lieu à une autre interprétation que celle dont découle la théorie motrice : l'invariant se situerait à un niveau syllabique, et non phonémique (Massaro, 1972).

1.3.2.1 Un débat alimenté par les paradigmes de détection de fragments

Les débats entre partisans du phonème et de la syllabe en perception ont été nourris notamment par des travaux sur l'identification d'unités, ou détection de fragments. Ce paradigme consiste à mesurer le temps de réaction à des tâches d'identification de stimuli auditifs. De nombreuses études ont soutenu le rôle de la syllabe comme unité en perception, en montrant que l'identification de syllabes était plus rapide que l'identification de phonèmes, en anglais et en français (Savin & Bever, 1970; Segui, Frauenfelder, & Mehler, 1981). Néanmoins, des défauts importants ont été révélés concernant la conception des stimuli utilisés dans la majorité de ces études, remettant en cause leurs conclusions (Norris & Cutler, 1988). En remédiant à ces défauts, Norris & Cutler (1988) ont observé que, en anglais, le phonème était identifié plus rapidement que la syllabe. En fait, les résultats d'études similaires menées dans d'autres langues, comme le japonais, ou l'espagnol, montrent que l'unité la plus rapidement identifiée dépend de la langue (Cutler & Otake, 2002; Otake, Hatano, Cutler, & Mehler, 1993).

D'autre part, un effet du statut lexical a été mis en évidence dans ce paradigme : les phonèmes ou les syllabes sont identifiés plus vite lorsqu'ils appartiennent à un mot qu'à un pseudo-mot (Rubin, Turvey, & Van Gelder, 1976). Et les mots entiers peuvent être identifiés plus rapidement que les syllabes (Foss & Swinney, 1973). Cela pourrait indiquer que l'unité de base de la perception est le mot, qui serait ensuite découpé en syllabe ou en phonème. Foss & Swinney (1973) proposent que la perception et l'identification sont deux processus distincts, avec chacun leur succession temporelle d'unités : la perception n'engendre pas nécessairement l'identification consciente.

Dans une étude en français, il a aussi été montré qu'une succession de phonèmes (ex : /bal/) est identifiée plus rapidement lorsqu'elle constitue une syllabe d'un mot (comme dans le mot 'balcon') que dans le cas contraire (comme dans 'balade', voir Mehler 1981). Cet effet a été répliqué en catalan et en hollandais (Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992; Zwitserlood, Schriefers, Lahiri, & van Donselaar, 1993), mais pas avec des anglophones, malgré de nombreuses tentatives (Cutler, Mehler, Norris, & Segui, 1983, 1986) ni dans aucune langue de type accentuée (Cutler, 1997). Ainsi, cela tend à confirmer que l'unité de base de la perception pourrait varier en fonction de la langue, et en particulier de sa structure rythmique. L'effet n'est pas non plus répliqué lorsque des pseudo-mots sont utilisés à la place de vrais mots (Content, Meunier, Kearns, & Frauenfelder, 2001).

D'autres méthodes expérimentales ont apporté leur contribution au débat. Le paradigme d'illusion perceptive est basé sur le mélange de deux pseudo-mots bisyllabiques, que le participant reçoit dans une oreille différente du casque (Kolinsky, Morais, & Cluytens, 1995). Ce qui est finalement perçu consciemment par le participant dépend de la répartition de l'information entre les deux oreilles. Le paradigme d'amorçage, déjà présenté pour l'étude des unités en production (voir section 1.2.1) est aussi utilisé en perception. Une étude en priming a par exemple suggéré que l'allophone jouait un rôle dans la perception (Luce et al., 2000). Ainsi, au-delà du débat phonème-syllabe, d'autres unités sont également proposées, telles que l'allophone, le trait phonétique (Lahiri & Reetz, 2010) ou encore le mot (Goldinger, 1988, voir aussi la section 1.2.2).

Néanmoins, la plupart des méthodes expérimentales évoquées ci-dessus sont critiquées. En effet, ces méthodes expérimentales, bien que destinées à explorer les unités de base de la perception, portent en fait bien souvent sur des unités associées à un traitement ultérieur de l'information. Ainsi, cette étape de traitement peut être la segmentation, l'encodage lexical, ou d'autres traitements post-perceptifs, voir méta-phonologiques (voir Kolinsky, 1998, pour une revue critique à ce sujet).

1.3.2.1 La généralisation pour révéler les unités

D'autres paradigmes sont proposés, avec pour but de mieux spécifier le niveau de traitement de l'information associé aux unités sondées (Mitterer, Reinisch, & McQueen, 2018). Les paradigmes d'adaptation sélective et d'apprentissage perceptif se basent tous deux sur le même principe (Bowers, Kazanina, & Andermane, 2016; Mitterer, Scharenborg, & McQueen, 2013). Il s'agit de provoquer un changement perceptif chez un individu, et d'observer la généralisation de ce changement à une autre situation. La généralisation est considérée comme le signe de l'existence d'une unité commune entre les deux situations.

Ces deux paradigmes se déroulent en deux temps. Le participant suit une phase d'adaptation (ou d'apprentissage) visant à décaler la frontière catégorielle entre deux phonèmes, par exemple /f/ et /s/, dans un contexte donné. Dans l'adaptation sélective, cette phase consiste en l'exposition passive à des stimuli contenant uniquement l'un des deux sons, par exemple /f/, avec « leaf » et « golf ». Dans l'apprentissage perceptif, cette phase est une tâche de décision lexicale, où par exemple le son /s/ est systématiquement remplacé par un stimulus ambigu. Ensuite, dans les deux paradigmes, une phase de test consiste à catégoriser des stimuli ambigus, en particulier des paires minimales ne différant que par les deux phonèmes ciblés /s/ ou /f/.

Dans les deux cas, la manipulation citée en exemple vise à augmenter le taux de réponse pour le mot contenant /s/. La phase de test peut permettre d'observer si l'apprentissage réalisé dans un contexte linguistique donné se généralise à un autre contexte. Il est supposé que la généralisation intervient lorsque la même unité perceptive est mise en jeu dans la phase d'entraînement et la phase de test.

Par exemple, si la manipulation appliquée dans la première phase ne concernait que le phonème placé en fin de mot, on peut observer si le changement de perception intervient aussi quand le même phonème est placé en début de mot. Ainsi, Mitterer, Scharenborg, & McQueen (2013), utilisant un paradigme d'apprentissage perceptif, n'ont pas observé de généralisation lorsque le phonème était placé dans une autre position, alors que l'apprentissage avait bien eu lieu. Ils en ont déduit que l'allophone jouait un rôle dans le système de perception. Néanmoins, Bowers et collègues (2016) ont remarqué que les stimuli ambigus construits pour cette étude étaient basés sur les signaux acoustiques d'allophones spécifiques, ce qui a pu influencer les résultats, et empêcher la généralisation au niveau du phonème. Dans une autre étude utilisant l'adaptation sélective, une généralisation a été observée pour les continua /f/-/s/ et /b/-/d/ d'une position à l'autre dans le mot, notamment début et fin de mot, donc d'un allophone à un autre (Bowers et al., 2016). Ces résultats soutiennent l'existence d'une représentation du phonème intervenant dans la perception de la parole. Une autre étude d'adaptation sélective a cependant apporté des éléments supplémentaires en faveur de l'allophone (Mitterer et al., 2018).

1.3.3 Conclusion

Si la production de parole utilise la perception (voir section 1.4), la perception repose aussi en partie sur la production. En effet, dans la plupart des théories de la perception de la parole, on retrouve une implication des représentations articulatoires ou motrices issues de la production de parole (Liberman et al., 1967). Les théories perceptuo-motrices proposent l'existence de représentations sensorimotrices, associant des caractéristiques acoustiques et motrices (Schwartz et al., 2012).

En parallèle, un autre débat a lieu concernant la nature des unités linguistiques à la base de la perception. La syllabe et le phonème y tiennent une place centrale, la première plutôt associée aux langues syllabiques et le second mis davantage en avant dans les langues accentuées. Le mot, le trait phonétique et l'allophone sont aussi défendus par certaines théories, reposant sur différents paradigmes expérimentaux. L'une des difficultés est de trouver une méthode expérimentale pointant spécifiquement sur les représentations intervenant en perception, et non

sur les représentations associées à d'autres processus de plus haut-niveau, comme l'identification consciente d'une unité (Kolinsky, 1998). Le transfert d'apprentissage perceptif semble répondre à ce problème (Mitterer et al., 2018). Ce paradigme utilise la généralisation (ou *transfer*) comme un moyen de sonder les représentations de la perception de parole.

1.4. Développement des unités de parole chez l'enfant

Dans les sections précédentes, nous avons pu constater que de nombreuses visions différentes s'opposent sur la question de la nature et du niveau de granularité des unités de parole chez l'adulte. Une des manières de les départager est de voir lesquelles sont compatibles avec le développement de l'enfant (Redford, 2015). En effet, pour comprendre la nature et la structure des représentations qui sous-tendent la parole, observer comment elles se construisent semble particulièrement indiqué.

Chez l'enfant, le développement de la perception de la parole a un temps d'avance sur la production, et les unités de base sous-tendant ces deux processus semblent se développer selon un chemin différent. On verra néanmoins que développement de la perception et de la production de parole sont intimement liés. Les grandes étapes du développement du langage chez l'enfant sont notamment synthétisées dans la revue de Kuhl (2004). Nous examinerons les principales étapes, en perception et en production, avec une attention particulière portée aux unités de base qui les sous-tendent.

La Figure 1.3 propose une représentation graphique des thèmes traités dans cette section. Dans la Figure 1.4, nous proposons également une représentation chronologique de ces thèmes et de ceux de la section suivante, pour aider le lecteur à se repérer. Seul le développement typique sera traité dans cette section. Pour un état de l'art complet sur le développement du contrôle moteur de la parole chez l'enfant, le lecteur pourra se référer à la thèse de Barbier (2016). La thèse de Barnaud (2018) apporte par ailleurs une revue de la littérature sur l'apprentissage statistique, et sur les modèles d'apprentissage en perception, qui ne sont pas développés dans ce manuscrit.

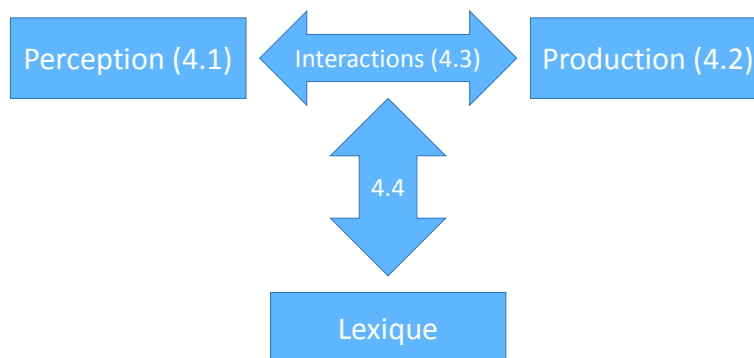


Figure 1-3: Schéma des thèmes abordés dans cette section consacrée au développement

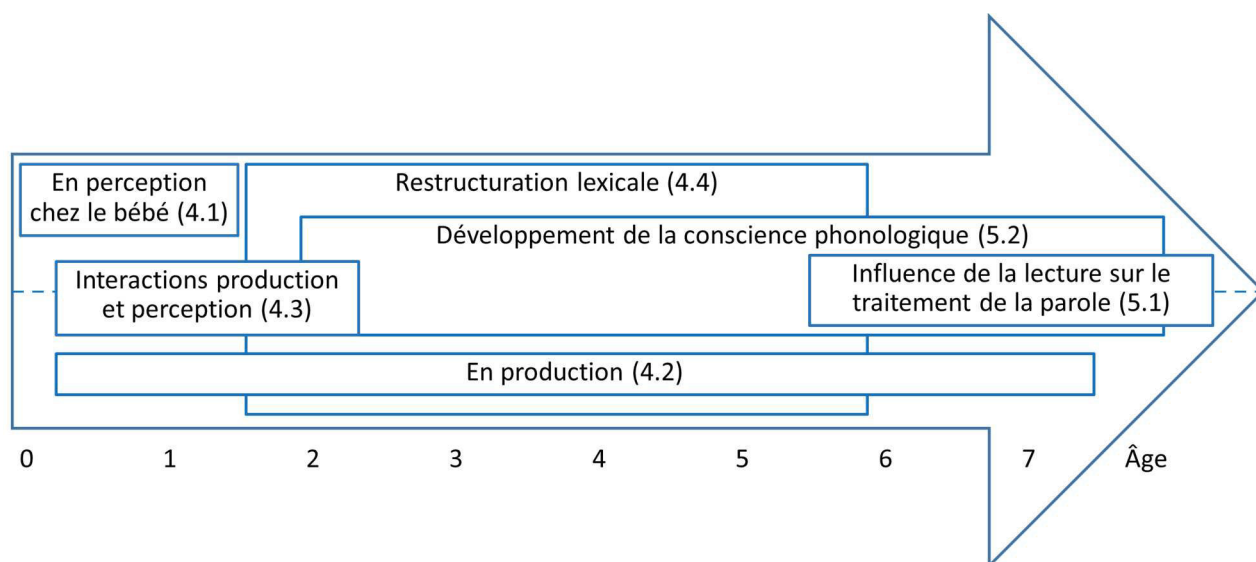


Figure 1-4: Frise chronologique situant chaque sous-section de la partie développementale de ce chapitre 1 en fonction de l'âge, et de si elle est en relation avec la perception de la parole (en haut), la production de parole (en bas) ou les deux.

1.4.1 En perception chez le bébé

L'enfant est capable dès les premiers mois de vie de distinguer des contrastes phonétiques. Il montre déjà une propension à la perception catégorielle, c'est-à-dire une meilleure discrimination de deux sons appartenant à deux catégories phonologiques différentes que de deux sons de même proximité acoustique, mais appartenant à la même catégorie phonologique. En réalité, chez l'adulte, la perception catégorielle désigne le cas extrême où on ne peut discriminer que deux sons de catégories différentes, et où la discrimination entre deux sons est entièrement prédictible par la catégorisation (Liberman, Harris, Hoffman, & Griffith, 1957).

Eimas, Siqueland, Jusczyk, & Vigorito (1971) ont montré que des nourrissons de 1 à 4 mois étaient capables, en ce sens, de distinguer les phonèmes /b/ et /p/ dans un contexte syllabique, c'est-à-dire en fait les syllabes /ba/ et /pa/. Ils ont utilisé pour cela une mesure de la vitesse de succion, qui est connue pour diminuer lorsque le bébé s'habitue à un stimulus répété, et qui connaît au contraire un regain immédiat lorsque le bébé est exposé à un stimulus qu'il perçoit comme nouveau. Cette expérience a été largement répliquée, pour divers types de contrastes phonétiques : dès 1 ou 2 mois, les nourrissons se montrent aptes à discriminer des contrastes phonétiques portant sur le lieu d'articulation (Eimas, 1974; Trehub & Rabinovitch, 1972), le mode d'articulation (Eimas & Miller, 1980), ou encore des contrastes concernant les voyelles (Grieser & Kuhl, 1989; Kuhl, 1991; Trehub, 1973, 1976).

Ces capacités intrinsèques sont ensuite peu à peu influencées par les propriétés du langage ambiant, et le fait qu'il soit associé ou non à du sens. Le bébé peut distinguer très tôt les contrastes phonétiques de sa langue maternelle, mais également des contrastes d'autres langues n'existant pas dans sa propre langue (Kuhl et al., 2006; Trehub, 1976). Il se spécialise par la suite dans les contrastes de sa langue, aux environs de 6 mois pour les voyelles et 9 à 12 mois pour les consonnes, acquérant une meilleure précision, au détriment de sa capacité de discrimination des contrastes des autres langues du monde, qu'il perd en grande partie : c'est ce qu'on appelle le *perceptual narrowing* (Kuhl et al., 2006; Werker & Tees, 1984). Pour réaliser ce type d'étude avec des enfants de cet âge-là, on utilise la méthode *visual reinforced head turn procedure* (HPP ou HT), basée sur l'idée que le bébé focalise davantage son attention (et donc son regard) vers les stimuli nouveaux que vers les stimuli auxquels il est déjà habitué.

La capacité de l'enfant à discriminer des contrastes interagit avec le niveau lexico-sémantique. Elle semble en effet diminuer vers l'âge de 14 mois, lorsque l'enfant produit ses premiers mots. En fait cette apparente régression a lieu dans le contexte spécifique où l'enfant est amené à associer un sens au son entendu. Ainsi l'apprentissage lexical s'accompagne d'une perte de précision au niveau phonétique en perception. Lorsque des conditions expérimentales ingénieuses ne favorisent pas l'association à un sens, (i.e. en affichant à l'écran un échiquier en fond à la place d'un objet) l'enfant conserve ses capacités de discrimination (Stager & Werker, 1997).

Enfin, même si des phénomènes perceptifs de type catégoriel sont mis en évidence très tôt chez l'enfant, les processus de catégorisation s'affinent jusqu'à l'âge adulte, puisque les courbes des frontières catégorielles, pour les voyelles comme les consonnes, évoluent jusqu'à

l'adolescence, devenant de plus en plus précises et abruptes avec l'âge (Hazan & Barrett, 2000; Walley & Flege, 1999).

La mise en évidence des capacités de perception du nourrisson a constitué un argument en faveur de visions innéistes et formalistes du langage. Ces visions considèrent entre autres que de nombreuses capacités langagières de l'enfant sont purement innées, et associées à des fondations biologiques propres à l'homme (Chomsky 1959, voir aussi la revue de Gervain & Mehler, 2010)). En particulier, il existerait des représentations phonologiques, notamment du phonème, dans le cerveau du bébé dès sa naissance. Nous verrons par la suite que d'autres visions, qualifiées de fonctionnalistes (voir Vihman, DePaolis, & Keren-Portnoy, 2009), d'émergentistes ou de constructivistes, proposent que les capacités langagières du bébé émergent en grande partie de capacités générales du bébé (non spécifiques au langage) en association avec son expérience sensorielle et motrice. Ces visions se basent notamment sur l'observation de la production de parole des enfants, et son interaction avec la perception.

1.4.2 En production : du babillage à l'âge adulte

En production, les vocalisations, ou proto-voyelles, apparaissent vers l'âge de 3 ou 4 mois. Elles s'appuient sur des mécanismes de production similaires à ceux des voyelles (vibrations des cordes vocales, sans génération de turbulences). La période de 6 à 12 mois est marquée par la phase de babillage, qui consiste en la production de syllabes de type CV (Consonne-Voyelle) selon un certain rythme. Ces vocalisations de type syllabique consistent en l'alternance de phases d'ouverture et de fermeture de la mandibule, sans contenu linguistique. Les syllabes sont d'abord répétées ('baba'), c'est le babillage canonique. Par la suite, les syllabes d'un même enchaînement sont variées ('pata'), ce qui correspond au babillage dit diversifié, ou *variegated babbling* (Fagan, 2009; B. L. Smith, Brown-Sweeney, & Stoel-Gammon, 1989).

Le babillage est probablement une exploration qui permettrait au bébé de faire le lien entre une commande motrice et ses conséquences sensorielles, en particulier acoustiques. Cette exploration n'est d'ailleurs pas spécifique à la parole. Au même âge, le bébé réalise des mouvements rythmiques et répétitifs avec d'autres parties de son corps. Ces mouvements permettent également au bébé d'apprendre à prévoir les conséquences sensorielles de ses commandes motrices. Ainsi, le babillage peut être englobé dans un processus de développement sensorimoteur plus large (Fagan, 2009; Iverson & Fagan, 2004; Thelen, 1981). Cette idée que le babillage est ancré dans un processus moteur non spécifique au langage est également présente dans la Théorie du Cadre puis Contenu (Frame then Content) de (MacNeilage & Davis,

2001; MacNeilage, Davis, Kinney, & Matyear, 2000; MacNeilage, Davis, & Matyear, 1997). Les séquences de syllabes du babillage y sont vues comme le résultat d'oscillations mandibulaires plutôt que la combinaison d'une consonne et d'une voyelle. Ces mouvements répétés de la mâchoire auraient lieu dans un premier temps sans être accompagnés par un contrôle des articulateurs de la bouche (langue et lèvres). Il existe en effet des combinaisons universelles privilégiées des lieux d'articulation de la consonne et de la voyelle associées (on parle de « co-occurrence », par exemple, consonne labiale et voyelle centrale, comme dans /ba/, et consonne dentale/alvéolaire et voyelle antérieure, comme dans /di/). Le contrôle des articulateurs de la bouche viendrait ensuite progressivement, se traduisant par l'émergence du babillage diversifié.

Bien qu'il repose sur des mécanismes sensorimoteurs probablement plus généraux, le babillage n'en joue pas moins un rôle essentiel dans la suite du développement de la parole. Il a une influence sur la perception, en focalisant l'attention du bébé sur les sons qu'il est capable de produire dans un premier temps, et par la suite sur ce qu'il n'est pas capable de produire (DePaolis et al., 2011; Majorano, Vihman, & DePaolis, 2013; Vihman, Depaolis, & Keren-Portnoy, 2014). Ainsi la production de l'enfant a une influence sur la perception, et est critique dans le développement des unités phonologiques. De plus, la quantité et la variété du babillage ont été associées à un accroissement plus rapide du vocabulaire et une meilleure mémoire phonologique quelques mois plus tard (Keren-Portnoy, Majorano, & Vihman, 2009; Keren-Portnoy, Vihman, DePaolis, Whitaker, & Williams, 2010). Ainsi, le babillage prépare la production des premiers mots.

L'enfant produit souvent ses premiers mots vers l'âge d'un an. Les premiers mots peuvent se caractériser par une certaine similarité phonétique avec le mot dans sa forme canonique, associé à une cohérence dans le contexte d'utilisation, qui suggère l'association à une signification, bien que cela ne soit pas évident à identifier (Vihman & McCune, 1994). Les premiers mots de l'enfant sont assez précis du point de vue de l'articulation, alors que les suivants sont souvent plus approximatifs (Ferguson & Farwell, 1975). C'est parce que l'enfant sélectionnerait d'abord les mots qu'il est capable de produire, grâce au *filtre articulatoire* construit avec le babillage, qui rend ces mots plus saillants perceptivement (Vihman et al., 2014). On retrouve souvent dans les premiers mots des enfants les mêmes *schémas* ou modèles phonétiques, désignés par le terme de *template* en anglais (traduit aussi en français par le terme « gabarit », voir Wauquier, 2006). Ces *schémas* sont spécifiques à chaque enfant, et sont basés en partie sur leur expérience

du babillage. De plus, ils ne sont pas parfaitement spécifiés phonétiquement, c'est-à-dire qu'ils sont définis de manière globale, ou holistique, et non de manière segmentale. La théorie élaborée par Vihman & Croft (2007) à ce sujet se base en grande partie sur l'observation et la transcription phonétique de productions d'enfants, suivis de manière longitudinale. Pour Vihman & Croft (2007), le *template* reste l'unité fondamentale de production de la parole, même à l'âge adulte, comme proposé par les théories de l'exemplaire, ou de la phonologie basée sur l'usage (Pierrehumbert, 2001). Redford (2014, 2016) soutient aussi cette idée, et propose que les *templates* (ou schémas) sont fusionnés (au fur et à mesure de notre expérience de la parole) pour constituer des unités encore plus larges, comme des expressions ou des phrases que l'on prononce souvent, toujours associées à un but fonctionnel de communication.

L'idée que l'enfant représente les premiers mots qu'il produit de manière holistique a été cependant contestée par des études en perception (Swingley, 2003; Swingley & Aslin, 2000, 2002). Ces études ont montré que la reconnaissance des mots était affectée par la substitution d'une seule consonne dans un mot chez les 12-24 mois et ne dépendait pas de l'âge, ni de la taille du vocabulaire. Néanmoins, Stager & Werker (1997) ont montré que cela n'est vrai qu'en l'absence d'association au sens (comme évoqué dans la sous-section précédente 4.1). En effet, les enfants de 14 mois sont capables de discrimination phonétique précise en l'absence de contexte lexical (s'ils ne cherchent pas à associer un mot à un objet ou un sens) mais ont bien un encodage plus global lorsqu'ils apprennent un nouveau mot.

Pour étudier la formation des représentations phonologiques en production, une possibilité est d'observer la stabilité ou la variabilité de production, en termes acoustiques ou éventuellement articulatoires. Des études acoustiques se sont intéressées à la formation des catégories phonémiques, en particulier à leur stabilité, et s'il y a recoupement entre elles ou pas. Elles montrent que les voyelles sont en place (stables et sans recoupement) dès l'âge de 2 ans et demi (McGowan, McGowan, Denny, & Nittrouer, 2014). Quant aux consonnes, les plosives sont souvent les premières acquises, vers l'âge de 18 mois, tandis que pour d'autres consonnes plus difficiles à prononcer, les fricatives, l'évolution se poursuit jusqu'à l'âge de 13 ans (Romeo & Hazan, 2013).

La précision acoustique ou articulatoire est un indice de la formation des catégories, mais ne peut pas constituer à elle seul une preuve de l'existence de l'unité ou des unités de base utilisées. En effet, on peut imaginer que des catégories phonologiques soient bien formées, mais que leurs réalisations motrices soient perturbées, ou peu précises, du fait de la non maturité motrice des

enfants (Barbier, 2016). La compréhension des interactions entre représentations phonologiques (ou unités de planification) et réalisation par le système moteur en production de parole est un enjeu central des travaux sur la coarticulation, en particulier chez l'enfant (Barbier, 2016). Il existe actuellement assez peu de travaux sur la coarticulation chez l'enfant, avec des résultats parfois contradictoires (Barbier et al., 2015; Noiray, Abakarova, Rubertus, Krüger, & Mark, 2018; Zharkova, Hewlett, & Hardcastle, 2012).

1.4.3 Interactions production et perception

Bien que la production et la perception aient été historiquement souvent étudiées séparément, il existe une influence mutuelle entre les deux. L'influence de la perception sur la production semble évidente : très tôt le bain de langage ambiant, en particulier la langue maternelle, a une influence sur les productions du bébé (Kuhl & Meltzoff, 1996). L'influence réciproque a aussi été mise en évidence tout au long du développement.

Déjà chez le nourrisson de 4 mois, la perception est influencée par la position et les mouvements de ses propres lèvres (Yeung & Werker, 2013). De plus, Bruderer, Danielson, Kandhadai, & Werker (2015) ont montré que le port d'un anneau de dentition contraignant la position de la langue du bébé de 6 mois affecte sa perception des sons. Ce résultat suggère qu'il utilise dès cet âge-là des informations somatosensorielles qui interviennent dans la perception de la parole, probablement issues de sa propre expérience de production (Vilain, Dole, Loevenbruck, Pascalis, & Schwartz, sous presse).

Ensuite, les syllabes produites lors du babillage exercent une influence sur l'attention portée par l'enfant à ce qu'il entend, comme cela a été évoqué précédemment (DePaolis et al., 2011; Majorano et al., 2013). De plus, plusieurs études chez l'adulte et chez l'enfant ont montré que le fait de prononcer un mot nouveau (ou un pseudo mot) à haute voix (ou même le murmurer) permet de mieux le fixer en mémoire que le lire en silence ou l'entendre, et facilite sa reconnaissance ultérieure (Icht & Mama, 2014; MacLeod, 2011; Ozubko & MacLeod, 2010). Cette observation a été désignée par le terme d'effet de production. Ainsi la production joue un rôle dans l'apprentissage lexical de l'enfant, et l'accroissement de son vocabulaire réceptif.

Au global, ces observations soutiennent l'idée que l'enfant perçoit la parole en partie en utilisant des informations qui viennent de son expérience de la production de parole (Yeung & Werker, 2013). Ainsi, l'unité de parole chez l'enfant se construit à l'intersection de la production et de la perception, d'où émerge sa nature sensorimotrice, que l'on retrouve chez l'adulte (Vilain, Dole, Loevenbruck, Pascalis, & Schwartz, sous presse).

1.4.4 Développement phonologique et lexical

La représentation holistique des premiers mots deviendrait ensuite mieux détaillée phonétiquement, avec l'expérience du langage. Cette spécification pourrait se faire par l'observation des similarités entre les mots, le besoin d'acquérir un vocabulaire varié exerçant une pression sur l'acquisition de l'inventaire phonétique (Cristia & Hallé, 2012). Plusieurs théories portent sur le développement des unités phonologiques au cours de l'enfance.

Le modèle de restructuration lexicale (ou *Lexical Restructuring Model*) propose que cette spécification des mots a lieu sous la pression de l'accroissement du vocabulaire de l'enfant (Walley, Metsala, & Garlock, 2003). Cette restructuration lexicale interviendrait mot par mot, selon divers facteurs : taille et vitesse d'expansion du vocabulaire, fréquence et familiarité des mots, et surtout nombre de mots proches d'un point de vue phonologique (ou densité du voisinage phonologique). Lorsqu'un enfant acquiert plusieurs mots voisins, en particulier des paires minimales (qui ne diffèrent que par un seul contraste phonémique), il a besoin d'encoder ces mots de manière précise pour pouvoir les distinguer. Cette proposition se base également sur des expériences utilisant un paradigme de *gating*, qui consiste en un dévoilement partiel du stimulus (Metsala, 1997). Dans ce paradigme, le participant entend une partie d'un mot, un peu plus longue à chaque présentation, et il doit identifier le mot dès que possible (Grosjean, 1980). Metsala (1997) a observé que le temps de réaction dépend de l'interaction entre la fréquence des mots et la densité phonologique du voisinage (avec l'identification la plus rapide pour les mots fréquents ayant une faible densité de voisins). Les performances dans cette tâche évoluent entre 7 et 11 ans, et même après. Il a aussi été montré que la précision de l'articulation d'un mot est liée à la densité phonologique de son voisinage immédiat, c'est-à-dire au nombre de mots voisins, chez des enfants de 2 ans (Sosa & Stoel-Gammon, 2012). En retour, la précision de l'encodage du mot facilite certainement l'acquisition rapide de nouveaux mots, donc la croissance du vocabulaire productif et du vocabulaire réceptif (Carroll, 2001). Ainsi, les développements phonologique et lexical sont intimement liés (Stoel-Gammon, 2011).

Cette restructuration vers des représentations plus détaillées se retrouve également dans l'acquisition de la conscience phonologique. L'enfant acquiert en effet une capacité à identifier, reconnaître et manipuler explicitement des unités phonologiques de plus en plus petites : la syllabe, l'attaque (*onset*) et la rime, et pour finir le phonème (Ziegler & Goswami, 2005, voir la section suivante pour plus de détails). En se basant notamment sur des travaux sur la conscience phonologique, la Psycholinguistic Grain Size Theory propose que la représentation

du phonème en tant que catégorie n'émerge vraiment qu'avec l'apprentissage de la lecture (Ziegler & Goswami, 2005). Pour d'autres, la représentation du phonème émerge en l'absence d'alphabétisation, bien que la conscience phonémique, elle, ne se développe qu'avec la lecture (Ventura, Kolinsky, & Fernandes, 2007). Une synthèse des principales théories du développement phonologique, en relation avec l'acquisition de compétences associées à la lecture, est proposée dans la thèse de Ainsworth (2015).

L'une des difficultés pour départager ces visions est de trouver des méthodes permettant de sonder les représentations phonologiques implicites, par opposition à la conscience phonologique, qui porte sur les représentations explicitement manipulables (Ainsworth, 2015). Ainsworth (2015) a proposé différentes méthodes en réception, c'est-à-dire ne faisant pas intervenir la production de parole. La plupart de ces méthodes sont basées sur des jugements de similarité (évaluée avec des tâches différentes), où l'on peut distinguer la proximité phonémique (nombre de phonèmes partagés) d'une proximité à un niveau de granularité plus global.

1.4.5 Conclusion

En résumé, l'enfant démontre des capacités de discrimination des unités de parole en perception, dès les premiers mois de vie. En particulier il discrimine des syllabes ne différant que par un seul phonème (Eimas et al., 1971). La syllabe constitue la première unité de production chez l'enfant lors de la phase de babillage, qui prépare la production des premiers mots. Ces premiers mots reposent sur la production de *templates*, ou schémas moteurs globaux, correspondant à un mot entier. Alors que la mise en évidence des capacités perceptives du bébé a favorisé une vision formaliste, l'étude des productions de l'enfant a soutenu le développement d'une vision émergentiste des unités.

Bien que les unités en perception et en production se développent dans les premières années de vie selon deux trajectoires très différentes, elles sont en étroites relations. Dans la suite du développement de l'enfant, les deux trajectoires semblent se rejoindre, et l'étude des unités phonologiques est moins cloisonnée entre perception et production. La recherche s'intéresse tout particulièrement aux représentations lexicales. A partir de l'âge de 3-4 ans, une notion largement employée pour caractériser le développement phonologique de l'enfant est la conscience phonologique. Ce développement phonologique est marqué par un événement majeur chez la plupart des enfants : l'apprentissage de la lecture, vers l'âge de 6 ans. Nous consacrons la section suivante entièrement à ces deux compétences, que sont conscience

phonologique et lecture, à leurs relations mutuelles, et à leur interaction avec le traitement phonologique de la parole.

1.5. Apprentissage de la lecture et développement de la conscience phonologique

Dans cette partie, nous allons montrer en quoi l'apprentissage de la lecture et de l'écriture interagit avec le traitement de la parole, et en particulier avec les représentations phonologiques. Quand cela n'est pas précisé, nous nous baserons implicitement sur un système d'écriture alphabétique, qui propose la relation la plus directe entre graphème et phonème (par opposition à d'autres systèmes d'écriture, comme celui des idéogrammes chinois). Nous nous intéresserons tout particulièrement à la conscience phonologique, qui constitue un marqueur du développement phonologique, et dont les relations avec l'apprentissage de la lecture ont été largement étudiées. Le lecteur pourra également trouver des éléments à ce propos dans la thèse d'Ainsworth (2015). La Figure 1-5 reprend le schéma de la Figure 1-3 et y intègre les thèmes abordés dans cette section.

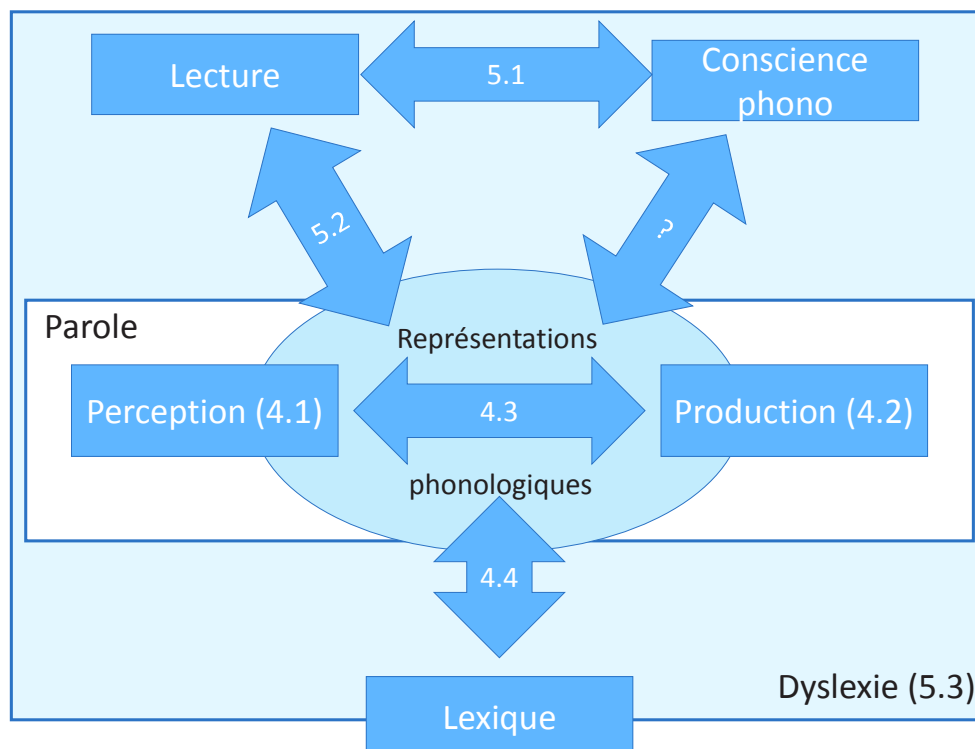


Figure 1-5: Schéma des thèmes abordés dans les deux sections consacrées au développement

1.5.1 La conscience phonologique

1.5.2.1 Conscience phonologique – définitions et évaluation

La conscience phonologique est une notion majeure dans le développement phonologique de l'enfant. En effet, à défaut de pouvoir sonder précisément les représentations phonologiques implicites de l'enfant, la recherche s'est intéressée à sa connaissance explicite des unités de parole.

Dans son sens le plus large, la conscience phonologique est la capacité à identifier et manipuler explicitement les unités phonologiques, telles que les syllabes, les rimes, ou les phonèmes. Certains distinguent la conscience épilinguistique, qui peut être définie comme une sensibilité générale à la similitude entre les sons, et la conscience métalinguistique correspondant à une conscience explicite des phonèmes, qui serait la « véritable » conscience phonologique (Gombert 1992). D'autres soutiennent que la conscience de la syllabe et de la rime sont des étapes vers la conscience du phonème, sans distinction de nature entre ces compétences, puisqu'elles sont corrélées (Anthony & Lonigan, 2004). L'âge d'apparition de ces différentes compétences peut néanmoins varier en fonction de la langue : par exemple, les enfants dont le français est la langue maternelle sont capables de distinguer et manipuler les syllabes plus tôt que les enfants ayant comme langue maternelle l'anglais (Duncan, Colé, Seymour, & Magnan, 2006). Dans ce travail de thèse, nous adoptons une définition de la conscience phonologique dans son sens le plus large, englobant la sensibilité à la rime et à la syllabe notamment. Lorsque nous parlerons spécifiquement du phonème, nous utiliserons le terme de conscience phonémique, correspondant au terme anglais *phonemic awareness*, couramment utilisé dans la littérature (voir par exemple Castles, Wilson, & Coltheart, 2011).

Il existe de nombreux types de tâches pour mesurer la conscience phonologique. Par exemple, reconnaître si deux mots riment (jugement de rimes), ou trouver l'intrus parmi une liste de mots qui riment sauf un (jugement de singularité). Il existe également des tâches analytiques, qui consistent à séparer les unités : compter le nombre de syllabes ou de sons d'un mot (ou plutôt taper avec la main à chaque syllabe ou son) ; ou enlever une syllabe (ou un son) d'un mot ; ou encore décomposer un mot en tous les sons qui le composent. D'autres tâches sont qualifiées de synthétiques : ajouter une syllabe ou un son à un mot ; ou bien mettre ensemble des sons ordonnés pour former un non-mot ou un mot (ou dire à quelle image correspond la suite de sons). Ainsi, la mesure de la conscience phonologique regroupe souvent plusieurs tâches à la fois. Des batteries de tests ont été développées dans différentes langues, pour différentes

tranches d'âge et usages (Lefebvre & Sutton, 2008). Il existe aussi des batteries de tests plus générales intégrant des tâches de conscience phonologique, comme le BELEC (Mousty, Leybaert, Alegria, Content, & Morais, 1999).

1.5.2.2 Conscience phonologique et apprentissage de la lecture

Les relations entre conscience phonologique et apprentissage de la lecture ont été étudiées abondamment. Alors que la *conscience syllabique*, ou capacité à identifier, reconnaître et manipuler les syllabes, présente dès l'âge de 3-4 ans, précède l'exposition au système alphabétique, la *conscience phonémique* semble apparaître simultanément à l'apprentissage de la lecture, et ce indépendamment de l'âge.

Il existe des études d'intervention, où l'on compare deux groupes d'enfants, l'un bénéficiant par exemple d'un entraînement à la conscience phonologique, quand l'autre groupe reçoit un entraînement dans un autre domaine. Les performances en lecture sont alors comparées. Il a ainsi été suggéré qu'un entraînement à la conscience phonologique peut faciliter l'apprentissage de la lecture (Treiman & Baron, 1983). Pour être efficace, l'entraînement à la conscience phonologique semble devoir être accompagné d'une connaissance de la correspondance entre sons et lettres (Castles & Coltheart, 2004; Castles, Coltheart, Wilson, Valpied, & Wedgwood, 2009; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012).

Par ailleurs, des études longitudinales ont permis de corréler développement de la conscience phonologique et apprentissage de la lecture, au fil du temps (voir Melby-Lervåg, Lyster, & Hulme, 2012, pour une méta-analyse). De ces études, il ressort un lien étroit entre conscience phonologique et lecture. Ce lien existe indépendamment d'autres facteurs contrôlés : milieu socio-économique, QI de l'enfant, et même vocabulaire et mémoire (Wagner et al., 1997). De plus, la conscience phonémique apparaît comme un meilleur prédicteur que la conscience d'unités plus grandes (Hulme et al., 2002). Pour certains, la conscience phonémique joue un rôle causal dans l'apprentissage de la lecture, qu'elle facilite (Melby-Lervåg et al., 2012). Ce rôle causal a été contesté par d'autres (pour une revue, voir Castles & Coltheart, 2004). En effet, la lecture et en particulier la connaissance de la correspondance entre sons et lettres, peut permettre d'améliorer les performances en conscience phonémique, simplement en appliquant des stratégies basées sur l'orthographe. L'influence serait donc bi-directionnelle. Par ailleurs, une étude menée dans 5 langues suggère que ce lien dépend de la langue étudiée, ce qui peut s'expliquer par une influence de la transparence orthographique, mais aussi par des pratiques d'enseignement différentes (Landerl et al., 2018). A l'inverse, d'autres études à travers

plusieurs langues soutiennent une conception plus universelle des liens entre conscience phonologique et lecture (Caravolas et al., 2012).

Enfin, les adultes illettrés ne sont pas capables d'ajouter ou d'enlever un son au début d'un pseudo-mot, contrairement aux adultes lettrés (Morais, Cary, Alegria, & Bertelson, 1979). Pourtant les adultes illettrés n'éprouvent pas de difficulté particulière pour la manipulation d'unités plus larges, telles que les syllabes et les rimes (Adrián, Alegria, & Morais, 1995; Bertelson, de Gelder, Tfouni, Morais, & Bertelson, 1989; Morais, Bertelson, Cary, & Alegria, 1986). Ce constat d'une absence de conscience phonémique a également été posé chez les lecteurs d'écritures non-alphabétiques, comme le Mandarin (de Gelder, Vroomen, & Bertelson, 1993; Read, Zhang, Nie, & Ding, 1986).

Ces observations sur la conscience phonologique invitent à se demander si la lecture opère une restructuration profonde des représentations phonologiques, et des traitements de la parole qui reposent sur elles. On peut en particulier se demander si l'acquisition de la lecture est à l'origine de l'émergence de la représentation mentale du phonème (Ziegler & Goswami, 2005). Pour tenter de répondre à cette question, nous allons examiner les interactions mises en évidence entre la lecture et le traitement de la parole.

1.5.2 Interactions entre la lecture et le traitement de la parole

Différents types de travaux ont mis en évidence une influence de la lecture (et de l'écriture) sur le traitement de la parole : des études comparant adultes illettrés et lettrés (ou moins couramment des locuteurs de langues non alphabétiques), et des études sur l'influence de l'orthographe des mots pour différentes tâches. Réciproquement, nous nous intéresserons dans un 2^{ème} temps aux précurseurs de la lecture, c'est-à-dire les compétences du langage oral qui semblent favoriser l'apprentissage de la lecture.

1.5.2.1 Effets orthographiques sur la parole

Tout d'abord, l'orthographe des mots influence leur reconnaissance auditive. Dans une étude pionnière sur l'influence de l'orthographe, Seidenberg & Tanenhaus (1979) ont montré que lors d'une tâche de reconnaissance de rimes, la similarité orthographique des mots avait un effet sur le temps de réaction des participants. L'influence orthographique a été mise en évidence dans de nombreuses tâches méta-phonologiques, comme le comptage de phonèmes (Treiman & Cassar, 1997; voir aussi Ventura, Kolinsky, Brito-Mendes, & Morais, 2001). Cependant, les tâches méta-phonologiques peuvent être en général réalisées avec des stratégies orthographiques (Goswami, 2002). Le participant peut en effet se baser sur la similarité

orthographique pour déterminer si deux mots riment ou non, sans utiliser leurs représentations phonologiques. Cela constitue une limitation importante de ces expériences, l'influence orthographique observée n'étant pas forcément reliée à un traitement phonologique.

Néanmoins, l'influence orthographique s'observe dans d'autres tâches sans qu'une stratégie orthographique puisse y être employée. Par exemple, la tâche de décision lexicale auditive consiste à déterminer si un stimulus entendu est un vrai mot ou pas. Tandis que la densité de voisins phonologiques a tendance à réduire les performances pour des tâches de décision lexicale, la densité de voisins orthographiques a l'effet inverse : plus un mot a de voisins orthographiques, plus il sera reconnu facilement (Ziegler, Muneaux, & Grainger, 2003). Il a été montré que ces effets de facilitation orthographique reposent sur le fait que l'information orthographique apporte une meilleure spécification des représentations phonologiques (Muneaux & Ziegler, 2004). Perre, Pattamadilok, Montant, & Ziegler (2009) ont également montré que les effets de cohérence orthographique en reconnaissance de mot étaient localisés dans une aire cérébrale associée au traitement phonologique (gyrus supramarginal gauche, dit SMG gauche, BA40), et sans activation concomitante particulière de l'aire de reconnaissance visuelle des mots (en anglais *Visual Word Form Area*, dite VWFA, dans le gyrus fusiforme gauche, BA37). Ces résultats soutiennent l'idée que l'orthographe des mots n'exerce pas qu'une influence momentanée liée à l'activation des formes orthographiques au cours de la tâche. Au contraire, l'orthographe semble jouer un rôle vraiment structurant dans les processus et les représentations phonologiques (Perre, Bertrand, Ziegler, Davis, & Holloway, 2011). De même, Pattamadilok, Knierim, Kawabata Duncan, & Devlin (2010) ont montré que l'inhibition de l'aire cérébrale associée à l'orthographe (cortex occipito-temporal ventral, vOTC) par stimulation transcrânienne magnétique (Transcranial Magnetic Stimulation, TMS) n'avait pas d'influence sur l'effet de cohérence orthographique, alors que la stimulation de l'aire de traitement phonologique (SMG) annulait cet effet. Les auteurs de ces travaux concluent que l'orthographe structure le traitement phonologique de la parole.

Cette restructuration des représentations phonologiques par l'orthographe (au-delà d'une possible activation momentanée par un stimulus visuel) est aussi suggérée par une autre étude, en production (Bürki, Spinelli, & Gaskell, 2012)). Cette étude a mis en évidence l'influence de l'orthographe sur la production et la reconnaissance des variantes phonologiques, c'est-à-dire de formes phonétiques non canoniques apparaissant parfois dans le langage courant. Cette influence a été montrée suite à une unique exposition orthographique, sur des pseudo-mots

appris dans le cadre de l'étude. Les résultats de l'étude suggèrent que des représentations de variantes phonologiques peuvent être créées sur la seule base de l'orthographe des mots, et donc que celle-ci joue un rôle structurant dans les représentations phonologiques, y compris en production de parole.

Quelques travaux se sont en outre intéressés à des données acoustiques et articulatoires pour mettre en évidence une influence de la lecture et de l'orthographe sur la production de parole (voir Saletta, 2015, pour une revue). Saletta, Goffman, & Brentari (2015) ont évalué la stabilité articulatoire (en se basant sur un indice d'ouverture des lèvres) et la précision segmentale (proportion de consonnes prononcées correctement) chez des adultes apprenant des pseudo-mots, en faisant varier la modalité de présentation (stimulus auditif ou mot écrit) et la transparence orthographique. Ils ont observé une meilleure précision articulatoire lorsque le pseudo-mot a été appris avec la modalité visuelle (mot écrit) plutôt qu'avec la modalité auditive, même lorsque le mot écrit n'était plus présenté. La précision segmentale était aussi meilleure avec un apprentissage dans la modalité visuelle, seulement dans le post-test. Ainsi, chez l'adulte, apprendre un mot nouveau en le lisant permet d'augmenter la stabilité articulatoire et la précision acoustique par rapport à seulement l'entendre et le répéter.

1.5.2.2 Effet de l'alphabétisation : études comparant populations lettrées et illettrées

Les populations d'adultes illettrés étudiées le sont en général pour des raisons sociales (et non pour des problèmes neurologiques par exemple) ce qui permet d'éliminer d'autres facteurs ayant pu avoir une influence sur leur traitement de la parole. Par exemple, dans certaines régions au Portugal il était courant de garder l'aîné à la maison pour les tâches domestiques et la prise en charge des enfants les plus petits, tandis que les enfants suivants allaient à l'école quand ils étaient en âge d'y aller. Ainsi ces populations d'adultes lettrés et illettrés sont comparables en de nombreux points, excepté leur exposition à l'instruction et en particulier à l'apprentissage de la lecture. Les études montrent des différences dans ces populations, sur des performances à certaines tâches, mais aussi concernant l'activation cérébrale associée, et même sur des aspects anatomiques du cerveau (voir Castro-Caldas & Reis, 2003, pour une revue des études sur les populations d'adultes illettrés).

En perception, Serniclaes, Ventura, Morais, & Kolinsky (2005) ont montré que les adultes illettrés disposaient d'une perception catégorielle mais que leur capacité de catégorisation était moins nette que chez les lettrés (frontières catégorielles moins précises), et associée à un biais lexical plus important (voir Ganong, 1980, pour l'expérience pionnière sur le biais lexical). Les

frontières moins nettes et le biais lexical plus fort sont d'ailleurs également observés chez les dyslexiques (voir Noordenbos & Serniclaes, 2015, pour une méta-analyse).

Les adultes illettrés ont notamment des difficultés à répéter des pseudo-mots (Reis & Castro-Caldas, 1997) alors qu'ils n'ont pas plus de difficultés que les autres à répéter de vrais mots. L'observation par imagerie fonctionnelle lors d'une tâche de répétition de mots et non-mots montre que l'apprentissage de la lecture a une influence sur l'architecture fonctionnelle du cerveau à l'âge adulte (A Castro-Caldas, Petersson, Reis, Stone-Elander, & Ingvar, 1998). En effet, la répétition de mots fait appel à un réseau neuronal plutôt lexico-sémantique, alors que la répétition de non-mots se base davantage sur les connaissances phonologiques, qui semblent reliées aux connaissances orthographiques chez l'adulte lettré, par contraste avec l'adulte illettré. De nombreuses études en neurosciences suggèrent que l'activation de certaines aires cérébrales lors du traitement de la parole dépend de l'alphabétisation (voir Dehaene, Cohen, Morais, & Kolinsky, 2015, pour une revue). Récemment, un effet de l'alphabétisation a aussi été montré sur la synchronisation des ondes cérébrales et la répartition de l'énergie dans les différentes bandes de fréquence (Araújo, Flanagan, Castro-caldas, & Goswami, 2018). Enfin, une étude en électrophysiologie utilisant les potentiels évoqués de réponse (ERP) de Wang, Wong, Wang, & Chen (2017) suggère que l'unité de planification principale de production de la parole serait la syllabe chez les locuteurs du Mandarin, contrairement aux langues germaniques. Cette étude suggère un rôle du système d'écriture (alphabétique ou non) dans les représentations mentales qui sous-tendent la production de parole.

En revanche, il existe aussi des études qui ne mettent pas en évidence de différence dans la représentation phonologique entre adulte lettrés et illettrés. Celles-ci sont moins nombreuses, mais il est difficile de savoir si cela est le reflet d'une réalité ou simplement d'un biais de publications favorisant les études montrant un effet. Par exemple, Ventura, Kolinsky, Fernandes, Querido, & Morais (2007) n'ont pas trouvé de différence entre adultes lettrés et illettrés sur une tâche de *gating* (voir section 1.4.2). L'étude suggère que la restructuration lexicale en représentations plus détaillées a lieu même en l'absence d'une exposition au système alphabétique, simplement comme conséquence du langage oral.

1.5.2.3 Les précurseurs de la lecture

Pour comprendre les relations entre lecture et parole, il peut être utile de s'intéresser à une littérature qui a regardé le problème sous un autre angle que celui adopté dans la section précédente : Qu'est-ce qui peut préparer le terrain, ou faciliter l'apprentissage, de la lecture

chez les enfants ? Cette question éclaire aussi les relations entre lecture et développement phonologique, et elle a été abondamment étudiée. En effet, cette question comporte un enjeu éducatif très important. Elle est aussi associée à un enjeu clinique de taille, celui de comprendre les sources des troubles d'apprentissage de la lecture, ou dyslexie.

La littérature sur la lecture et celle sur la dyslexie ont exploré de nombreux facteurs facilitant ou reliés à l'apprentissage de la lecture, notamment au travers d'études longitudinales. Les performances en lecture sont souvent évaluées à travers la lecture de non-mots, représentative du déchiffrage (Share, 1995), mais aussi par la lecture de mots, ou de textes (en fonction de l'âge des participants). Parmi les facteurs corrélés aux capacités de lecture les plus fréquemment étudiés, on trouve la conscience phonologique (définie précédemment), la mémoire de travail (ou mémoire à court-terme), la connaissance des lettres, la taille du vocabulaire (en production et/ou en perception), la performance à une tâche de dénomination rapide automatique (*Rapid Automatized Naming*, dit RAN). Cette tâche consiste à nommer des stimuli visuels ou des images familières (lettres, chiffres objets, ou couleurs par exemple) correctement et le plus rapidement possible (Araújo, Reis, & Petersson, 2015). Elle est reconnue comme une mesure de traitement phonologique implicite : elle repose sur des processus phonologiques, sans faire appel à une réflexion ou une conscience explicite à leur sujet (en particulier au sujet des unités phonologiques). D'autres mesures sont couramment utilisées. L'une d'elles est la mesure de l'empan numérique (dit *digit span* en anglais, ou mémoire des chiffres), qui révèle la capacité de mémoire à court terme, en particulier les performances de la boucle phonologique qui correspond au sous-système de la mémoire de travail qui repose sur la parole (Baddeley, 1983). En effet, lorsque l'on a besoin de retenir un numéro de téléphone par exemple, on se le répète dans la tête ou à haute voix. L'aptitude à se rappeler le numéro de téléphone dépend de la capacité de la boucle phonologique. Cette dernière serait étroitement liée au développement phonologique, plus précisément à la qualité des représentations phonologiques, sur lesquelles elle repose. Ainsi, pour certains, la mesure de performance de la boucle phonologique de la mémoire de travail et la mesure de la conscience phonologique font appel à une seule et même compétence (Wagner et al 1993). D'autres proposent que les tâches de mémoire de travail et la conscience phonologique reposent sur des mécanismes phonologiques communs, mais aussi sur des compétences distinctes (Alloway, Gathercole, Willis, & Adams, 2004).

Pour étudier les compétences phonologiques, on trouve aussi couramment des tâches de répétition de non-mots, ou de phrases, et d'acquisition de nouveaux mots. Ces tâches reposent

également fortement sur les capacités de la boucle phonologique. De plus, la capacité à acquérir de nouveaux mots est reliée à la taille du vocabulaire. Ainsi, la plupart des facteurs et mesures utilisés comme prédicteurs potentiels de la lecture sont corrélés entre eux. Un enjeu est alors de savoir s'il existe une seule ou plusieurs compétences indépendantes, et s'il y en a plusieurs, lesquelles permettent de prédire le mieux les performances futures en lecture (Shapiro, Carroll, & Solity, 2013). Il semble que la dénomination rapide automatique (RAN) et la conscience phonologique sont les deux mesures qui prédisent le mieux les compétences futures en lecture, selon des composantes indépendantes (Landerl et al., 2018).

Cependant, il existe peu d'étude à notre connaissance concernant d'éventuels facteurs sensorimoteurs servant de prédicteur pour l'acquisition de la lecture chez les enfants au développement typique. De tels facteurs sont néanmoins examinés dans la littérature sur la dyslexie développementale.

1.5.3 Dyslexie, représentations phonologiques et déficits sensorimoteurs

1.5.3.1 *Troubles associés à la dyslexie*

La dyslexie développementale (que nous désignerons par la suite simplement par dyslexie) est un trouble de l'apprentissage de la lecture qui touche environ 7% des gens, indépendamment de leur langue maternelle (Goswami, 2015). Elle est caractérisée par d'importantes difficultés de lecture et d'orthographe, qui persistent malgré une intelligence typique et une instruction habituelle (Boets et al., 2011; Boets, Wouters, Van Wieringen, & Ghesquière, 2006). La dyslexie est souvent associée à un déficit phonologique, qui constitue pour beaucoup le cœur et la source des difficultés d'apprentissage de la lecture (Ramus, 2004). Ce déficit phonologique a été mis en évidence à travers de nombreuses études sur la conscience phonologique chez les personnes avec dyslexie (voir Melby-Lervåg et al., 2012, pour une méta-analyse). Néanmoins, il semble qu'il existe différents sous-types de dyslexie, dont certains ne sont pas associés à un déficit phonologique, mais à un trouble visuo-attentionnel (Zoubrinetzky, Bielle, & Valdois, 2014; Zoubrinetzky, Collet, Serniclaes, Nguyen-Morel, & Valdois, 2016). La compréhension des causes de la dyslexie fait l'objet de vifs débats, entre tenants de diverses théories et hypothèses, parmi lesquelles les hypothèses des déficits phonologique, auditif, de traitement temporel (Tallal, 1980), du cervelet (Nicolson, Fawcett, & Dean, 2001), la théorie de la perception allophonique, ou encore la théorie magnocellulaire (qui vise notamment à unifier les déficits phonologiques et visuels, voir Stein 2001). Cette littérature est abondante, et les

relations entre ces théories sont complexes. Nous ne détaillerons donc pas chaque théorie, mais nous aborderons quelques travaux expérimentaux adoptant différents points de vue.

Si l'apprentissage de la lecture chez les enfants typiques a été relativement peu mis en relation avec des facteurs sensorimoteurs, aussi désignés par le terme de 'bas niveau', l'exploration de ces facteurs est beaucoup plus riche dans la littérature sur la dyslexie. Pour certains, la cause de la dyslexie serait d'abord de nature perceptive, et la mauvaise qualité des représentations phonologiques en serait une des conséquences. Pour d'autres, les troubles sensorimoteurs observés dans la dyslexie seraient plutôt une conséquence des problèmes de lecture ou des mauvaises représentations phonologiques (Ramus, 2004; White et al., 2006).

La dyslexie est associée à de faibles capacités de perception catégorielle (PC), avec des frontières catégorielles peu nettes (c'est-à-dire dont les courbes d'identification ont une pente maximale faible), et une faible discrimination des sons (voir Noordenbos & Serniclaes, 2015, pour une méta-analyse sur le sujet). Le constat de frontières catégorielles moins nettes a été confirmé récemment avec une méthodologie plus robuste que dans les études précédentes, éliminant de potentiels facteurs attentionnels (O'Brien, McCloy, Kubota, & Yeatman, 2018). Ce déficit en perception catégorielle engendre un déficit en conscience phonologique, dont découlent les difficultés en lecture : la conscience phonologique joue le rôle de médiateur dans ce trio (Zoubrinetzky et al., 2016).

La dyslexie étant souvent transmise de manière héréditaire, il est possible d'identifier des enfants à risque de dyslexie pour explorer de possibles marqueurs précoces de ce trouble développemental (Leppänen, Richardson, Eklund, & Guttorm, 2002). Des études en Electro-Encéphalo-Graphie (EEG) ont montré des différences permettant de prédire l'apparition de difficultés de lecture (Hämäläinen et al., 2013). Les capacités de détection de sons purs dans le bruit et de modulation de fréquence ont été corrélées positivement à la conscience phonologique, chez des enfants d'âge préscolaire (5 ans) à risque de dyslexie ou non (Boets, Wouters, van Wieringen, De Smedt, & Ghesquière, 2008). De plus, dans une étude longitudinale, la sensibilité aux modulations de fréquence ainsi que la perception de parole dans le bruit à l'âge de 5 ans ont permis de prédire la vitesse d'acquisition de la lecture, indépendamment de la performance en conscience phonologique (Boets et al., 2011). Ainsi, ces études suggèrent que les difficultés perceptives précèdent et sont à l'origine du déficit phonologique présent dans la dyslexie.

Par ailleurs, des problèmes d'équilibre, de stabilité de la posture, de coordination, de dextérité manuelle, et une certaine maladresse ont été mis en évidence chez les personnes dyslexiques (Nicolson & Fawcett, 1995; Ramus, Pidgeon, & Frith, 2003). Par exemple, les enfants et adolescents dyslexiques sont plus lents que des sujets-contrôles du même âge pour enfiler des perles (Fawcett & Nicolson, 1995). Des difficultés d'apprentissage moteur et d'automatisation ont aussi été révélées (Nicolson & Fawcett, 1990). Par exemple, les performances des adultes dyslexiques à une tâche de temps de réaction sériel (ou *Serial Reaction Time Task*, dite SRTT en anglais, voir Nissen & Bullemer, 1987) ne se sont pas améliorées sur des séquences répétées, au contraire des sujets contrôles (Stoodley, Harrison, & Stein, 2006). Il a également été suggéré que les personnes dyslexiques étaient plus lentes pour articuler les consonnes occlusives /p/, /t/ et /k/, qu'elles soient produites de manière isolée ou en séquence (Fawcett & Nicolson, 2002; voir aussi Malek, Gholizadeh, Amiri, Hekmati, & Pirzadeh, 2013). Pour certains, ces observations soutiennent l'hypothèse d'un déficit du cervelet qui serait à l'origine de toutes les difficultés observées dans la dyslexie, en particulier le déficit phonologique (Nicolson et al., 2001). Cette position est néanmoins contestée, notamment sur la base du fait que certaines aptitudes gérées par le cervelet, comme l'estimation du temps, ne font pas l'objet de difficultés particulières pour les dyslexiques (Ramus et al., 2003). De plus, les difficultés d'apprentissage implicites ont été remises en cause, avec une modification du paradigme du temps de réaction sériel (Kelly, Griffiths, & Frith, 2002). Un déficit attentionnel pourrait aussi être la cause de nombreuses difficultés observées (Ramus & Szenkovits, 2008).

1.5.3.2 *Intégration sensorimotrice et dyslexie*

Les problèmes de posture et d'équilibre évoqués précédemment pourraient aussi venir d'une différence au niveau de l'intégration sensorimotrice. Viana, Razuk, de Freitas, & Barela (2013) ont étudié les performances d'enfants dyslexiques et contrôles lors d'une tâche d'équilibre dans une pièce mouvante, selon qu'ils avaient les yeux bandés et ne pouvaient rien toucher, ou que leur vision et/ou leur retour somatosensoriel faisait l'objet d'une perturbation (pièce stationnaire mais barre mobile, ou inversement). Les enfants dyslexiques oscillaient plus que les sujets-contrôles dans toutes les conditions, et de manière moins cohérente lors des deux conditions où les modalités visuelles et somatosensorielles entraient en conflit. Cela pourrait venir d'un déficit d'intégration sensorimotrice (Viana et al., 2013).

L'intégration sensorimotrice est une composante de la perception de la parole, avec les modalités visuelle et auditive (Ito, Tiede, & Ostry, 2009). Elle est tout aussi essentielle à la

production de parole, qui fait intervenir les retours auditifs et somatosensoriels du locuteur (Houde & Nagarajan, 2011; Patri, Perrier, Schwartz, & Diard, 2018; Tourville & Guenther, 2011). Pour étudier l'intégration sensorimotrice en production de parole, il est possible de manipuler en temps réel le retour auditif ou somatosensoriel d'une personne en train de parler, comme cela sera détaillé dans le chapitre 2. On peut par exemple retarder le retour auditif (Yates, 1963), ou bien modifier la fréquence fondamentale de la voix (voir par exemple Scheerer & Jones, 2012). Il est même possible de modifier des paramètres du signal de parole entrant en jeu dans la catégorisation des unités phonologiques, comme les formants, ou fréquences des pics d'énergie du signal (Houde & Jordan, 1998, 2002). La perturbation des formants est donc une méthode toute indiquée pour comprendre les relations entre intégration sensorimotrice et représentations phonologiques.

Quelques études de perturbations de formants ont été menées récemment chez des personnes dyslexiques. Dans une étude de Van Den Bunt et al. (2017), les adultes dyslexiques se sont adaptés plus rapidement à la perturbation auditive (altérant /bep/ en direction de /bap/), lors de sa mise en place graduelle (rampe), et sont aussi moins bien revenus vers leur prononciation initiale lorsque le retour auditif revenait à la normale, en comparaison d'un groupe contrôle. Les résultats sont interprétés à la lumière de la théorie de l'attracteur perceptif (ou *perceptual magnet theory*), selon laquelle un prototype de chaque catégorie phonologique jouerait un rôle d'attracteur, qui rendrait peu discriminable les sons à son voisinage, et résulterait en une meilleure discrimination des sons aux alentours des frontières catégorielles (Feldman, Griffiths, & Morgan, 2009; Kuhl, 1991). Les personnes dyslexiques disposeraient ainsi d'un attracteur moins puissant que les autres (Van Den Bunt et al., 2017). Cela est en accord avec le déficit de perception catégorielle évoqué précédemment. Cette observation est aussi cohérente avec la théorie allophonique de la perception (Serniclaes, Heghe, Mousty, Carré, & Sprenger-Charolles, 2004) selon laquelle les dyslexiques auraient des catégories organisées davantage autour des allophones plutôt que des phonèmes.

Un retour à la normale plus faible après l'arrêt de la perturbation auditive a aussi été observé chez l'enfant, dans une autre étude (Van Den Bunt, Groen, van der Kleij, et al., 2018). De plus, un meilleur score de conscience phonologique était corrélé à une adaptation plus rapide, plus forte, et un retour plus lent à la normale. Dans une dernière étude, cette fois chez des enfants non dyslexiques, une plus forte adaptation a été trouvée chez les enfants lecteurs de 8 ans que

chez les enfants non lecteurs de 4 ans, et l'amplitude de l'adaptation était corrélée à la conscience phonologique des enfants de 4 ans (Van Den Bunt, Groen, Frost, et al., 2018).

En résumé, ces travaux indiquent que l'intégration sensorimotrice est différente chez les enfants dyslexiques et non dyslexiques. De plus, les études en parole montrent un lien entre l'intégration sensorimotrice et le développement de la lecture et de la conscience phonologique. Cela suggère que l'intégration sensorimotrice contribue à la formation des représentations phonologiques, et/ou qu'elles en sont une manifestation au niveau du comportement. Ces deux notions semblent ainsi en étroite interaction.

1.5.4 Conclusion

L'apprentissage de la lecture a une influence sur la parole et sur de nombreuses tâches de traitement phonologique. En particulier, elle est étroitement liée à la conscience phonémique. Cette aptitude à manipuler les sons de parole semble même conditionnée par l'exposition à un système alphabétique. Ainsi, on peut se demander si c'est l'apprentissage de la lecture qui crée les représentations mentales du phonème, ou si celles-ci se développent en son absence ou avant sa mise en place, avec le langage oral.

Bien qu'il existe un nombre important de tâches visant à évaluer le développement phonologique, ou la qualité des représentations phonologiques d'un individu, ces tâches font généralement appel à de multiples aptitudes, et ne permettent pas d'isoler les représentations phonologiques.

1.6. Bilan des méthodes d'investigation des unités et projet de cette thèse

La première section examinait la nature des unités en production de parole, à travers différents modèles et théories. Ces derniers ont tour à tour mis en avant le mot, la syllabe, le phonème, et le geste articulatoire comme unité de base de la production de parole. Dans la 2^{ème} section, nous avons mis en évidence la nature sensorimotrice des unités de parole. La section suivante s'intéressait à la construction des unités de parole chez l'enfant, et la dernière à l'influence de la lecture dans cette construction. Dans cette dernière partie du chapitre 1, nous nous focaliserons sur les méthodes d'investigation de la nature des unités de parole. La plupart de ces méthodes ont été mentionnées dans les parties précédentes. Nous en ferons ici une brève revue critique, qui nous amènera ensuite à exposer le projet de cette thèse.

Chez l'adulte, l'étude des unités de production s'est essentiellement basée sur les méthodes suivantes : analyse des erreurs de production, temps de réaction à des tâches de dénomination (ou de décision lexicale), ou encore paradigme de *priming* (ou amorçage). En perception, la littérature s'est beaucoup basée sur des tâches d'identification de fragment (syllabes ou phonèmes notamment). Par ailleurs, les paradigmes d'adaptation sélective et d'apprentissage perceptif se basent sur la notion de généralisation pour sonder le niveau de granularité des unités perceptives. Chez le bébé, l'étude des unités en perception a porté essentiellement sur ses capacités de discriminations des unités. En production, la recherche s'est intéressée aux caractéristiques phonétiques des mots produits (template), et à leur précision segmentale notamment. Le paradigme de *gating* (ou dévoilement partiel du stimulus) est utilisé pour sonder les unités qui forment le lexique. La conscience phonologique est une façon de mesurer le développement phonologique à partir de l'âge de 3 ans environ. Elle est un précurseur de la lecture, tout comme d'autres mesures : boucle phonologique, RAN (dénomination rapide automatique).

Des critiques ont été adressées à une partie de ces méthodes, principalement en perception. Si la plupart des articles utilisant ces méthodes ont pour ambition de mesurer « l'unité de base » de la perception ou de la production, des revues critiques ont pointé du doigt la difficulté de sonder de manière isolée les unités associées à une étape précise du traitement de la parole, telle que l'interface entre encodage phonologique et articulation, que ce soit en perception ou en production (Ainsworth, 2015; Kolinsky, 1998). Souvent les méthodes employées testent plusieurs étapes du traitement de la parole en même temps, sans pouvoir les dissocier. Ainsi, la conscience phonologique fait intervenir des traitements métalinguistiques de haut niveau. Dans sa thèse, Ainsworth (2015) a aussi fait le constat du manque de moyens pour évaluer les représentations phonologiques implicites, c'est-à-dire non accessibles à des processus conscients. Elle a proposé un ensemble de méthodes, essentiellement basées sur la perception.

Des critiques semblables pourraient être formulées pour les méthodes d'investigation des unités en production. Par exemple, l'analyse des erreurs de parole peut indiquer que telle ou telle unité linguistique intervient dans le processus de production d'une phrase. Mais elle ne permet pas d'indiquer si une unité est impliquée directement dans l'articulation de la parole, ou si elle intervient à une étape antérieure du processus d'encodage phonologique. L'observation d'effets de fréquence des unités sur le temps de réaction en production connaît la même limitation.

Par ailleurs, de nombreux travaux ont mis en évidence les interactions entre perception et production de la parole, et la nature sensorimotrice des unités de parole (Schwartz et al., 2012; Skipper et al., 2017; Yeung & Werker, 2013). Les modèles de contrôle moteur de la parole se basent en outre sur une correspondance entre articulation et conséquences perceptives. Notre thèse s'inscrit dans ce cadre qui considère que les unités assurant l'articulation de la parole sont sensorimotrices.

Ainsi, pour sonder les unités de base de la production de parole, l'enjeu est de trouver une méthode expérimentale permettant (1) de sonder sélectivement les unités sensorimotrices, et (2) d'en déterminer le niveau de granularité linguistique. Une méthode permettant de sonder sélectivement les unités sensorimotrices a été évoquée dans la sous-section concernant la dyslexie développementale : il s'agit du paradigme d'adaptation auditorimotrice (ou apprentissage auditorimoteur) en réponse à des perturbations de formants (Van Den Bunt, Groen, Frost, et al., 2018; Van Den Bunt et al., 2017; Van Den Bunt, Groen, van der Kleij, et al., 2018). Ce paradigme consiste à manipuler la relation entre les caractéristiques motrices et perceptives des représentations sensorimotrices, générant une forme d'apprentissage. D'autre part, la notion de généralisation, ou transfert, permet d'explorer le niveau de granularité de représentations faisant l'objet d'une adaptation, ou apprentissage. L'étude de la généralisation a notamment été proposée pour étudier le niveau de granularité des unités perceptives, en se basant sur un apprentissage perceptif, comme évoqué dans la section 1.3.2 (Bowers et al., 2016; Mitterer et al., 2013). Du couplage entre l'adaptation à des perturbations de formants et la notion de généralisation résulte le paradigme de transfert d'apprentissage auditorimoteur. Le transfert d'apprentissage auditorimoteur sera plus particulièrement détaillé dans une sous-section du chapitre 2, chapitre consacré à une revue de littérature sur le paradigme d'adaptation à des perturbations de formants.

L'objet de ce travail de thèse est de déterminer, grâce au paradigme de transfert d'apprentissage auditorimoteur, le niveau de granularité des représentations sensorimotrices qui sous-tendent l'articulation de la parole. Jusque-là, ce paradigme n'avait pas été directement exploité dans ce but, à notre connaissance. Nous nous intéresserons au niveau de granularité des représentations sensorimotrices chez l'adulte, et au cours du développement chez l'enfant, en particulier autour d'une étape majeure de restructuration phonologique : l'acquisition de la lecture.

2. Adaptation auditorimotrice : les perturbations de formants

2.1 Introduction

Dans le chapitre précédent, nous avons exploré au travers des principales théories et modèles de la littérature la nature sur les unités de production de la parole. Nous avons vu que le niveau de granularité de ces unités, chez l'adulte et au cours du développement de l'enfant, fait l'objet de débats toujours d'actualité, en psycholinguistique et au-delà. Par ailleurs, des travaux indiquent que les unités qui sous-tendent la parole sont de nature sensorimotrice, associant des caractéristiques motrices et perceptives. Enfin, à notre connaissance, aucune méthode destinée à déterminer le niveau de granularité des unités qui assurent l'articulation de la parole n'a exploité cette nature sensorimotrice.

Nous nous appuyons dans cette thèse sur un paradigme issu de la littérature du contrôle moteur pour répondre à cet enjeu : le transfert d'apprentissage sensorimoteur. Le raisonnement suivi repose sur l'idée suivante : si un apprentissage réalisé pour une tâche ou une situation donnée se généralise à une autre tâche (ou situation), c'est que la représentation mentale sur laquelle a porté l'apprentissage est commune aux deux tâches ou situations (respectivement l'apprentissage et le test). Ainsi le transfert d'apprentissage constitue une fenêtre comportementale pour sonder les mécanismes cognitifs sous-jacents à nos actions (Shadmehr, 2012) et en particulier, le niveau de globalité ou de spécificité des représentations impliquées dans ces actions.

Pour générer un apprentissage en conditions expérimentales, on utilise en général une perturbation, qui altère le mouvement et/ou sa perception par l'auteur du mouvement. Par exemple, l'atteinte d'une cible avec la main peut être perturbée par l'application d'une force appliquée sur une poignée tenue par la main (Mattar & Ostry, 2007, voir Figure 2.1(A)). L'apprentissage consiste à atteindre la cible de la manière la plus rapide et précise possible malgré la force appliquée. On peut alors tester la généralisation de cet apprentissage pour l'atteinte d'autres cibles situées ailleurs dans l'espace. Le transfert d'apprentissage a été également utilisé en perception de la parole pour sonder les unités perceptives (Bowers et al., 2016; Mitterer et al., 2013).

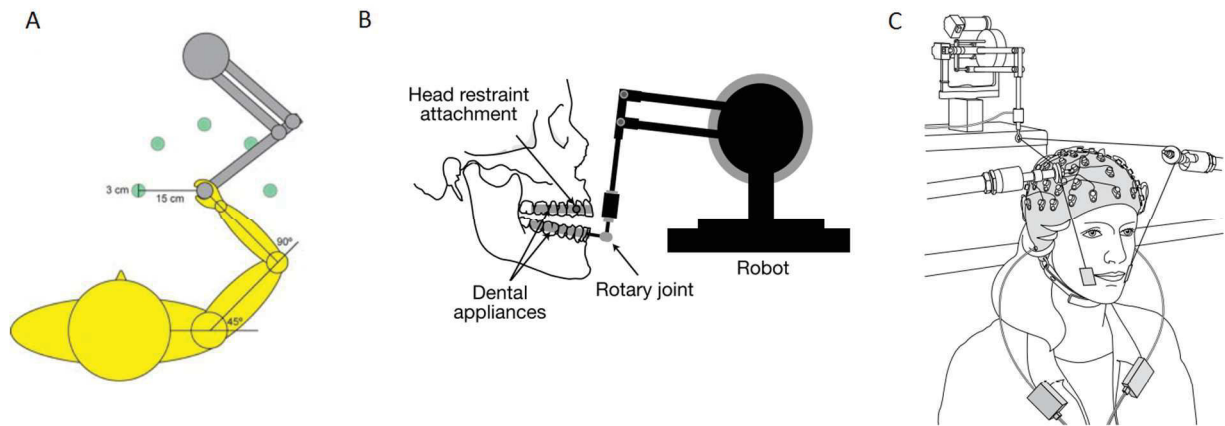


Figure 2-1: schéma de perturbations dynamiques (A) du geste du bras (Mattar & Ostry, 2007); (B) de la trajectoire de la mâchoire (Tremblay & al., 2003) et (C) du retour somatosensoriel (Ito & al., 2016). Ces figures sont extraites des articles correspondants.

Pour générer un apprentissage en production de parole, on peut utiliser différents types de perturbations. La pose d'un tube labial, dit *lip tube* en anglais, peut altérer la géométrie du conduit vocal du locuteur (Ménard, Perrier, Aubin, Savariaux, & Thibeault, 2008; Savariaux, Perrier, & Orliaguet, 1995). L'utilisation d'un système robotique (Phantom) permet d'appliquer une perturbation dynamique de la trajectoire de la mâchoire (Tremblay, Shiller, & Ostry, 2003, voir Figure 2.1(B)). Il existe aussi des perturbations purement perceptives, soit du retour somatosensoriel, par exemple en tirant sur la peau des joues (Ito, Coppola, & Ostry, 2016, voir Figure 2.1(C)), soit du retour auditif. Nous nous intéressons plus particulièrement aux perturbations auditives, qui s'obtiennent en décalant la fréquence fondamentale de la voix, ou bien les formants. Les formants sont les fréquences des pics d'intensité du signal acoustique qui traduisent le filtrage effectué par le conduit vocal, comme indiqué sur la Figure 2-2(A), et qui informent donc sur la position des articulateurs (mâchoire, langue, lèvres). Ce sont les formants qui permettent de distinguer la plupart des phonèmes (en particulier les voyelles) dans le domaine acoustique, représenté sur la Figure 2-2(B). Ainsi ce type de perturbation est particulièrement approprié pour l'étude des unités phonologiques.

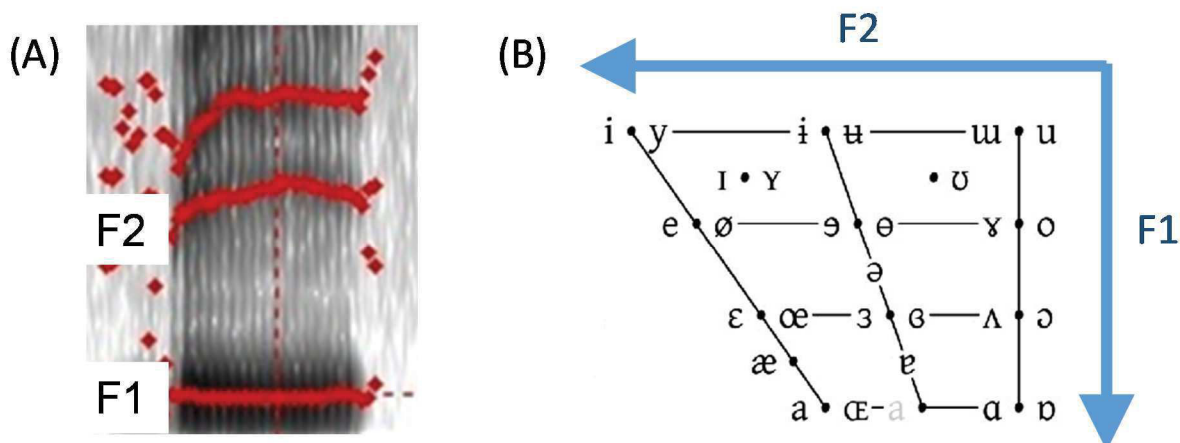


Figure 2-2: Les formants. (A) Spectrogramme (représentation avec le temps en abscisse et la fréquence en ordonnée) d'une syllabe où les formants F1 et F2 sont surlignés. (B) Trapèze vocalique dans l'espace (F1, F2)

Ce chapitre propose une revue de littérature sur les études utilisant des perturbations de formants. Il correspond à une publication pour un ouvrage sur l'apprentissage et la mémoire dans la perception et la production de la parole qui paraîtra prochainement:

Caudrelier, T., & Rochet-Capellan, A. (sous presse). Changes in speech production in response to formant perturbations: An overview of two decades of research. In S. Fuchs, J. Clelland, & A. Rochet-Capellan (Eds.), *Speech production and perception: Learning and Memory*. Peter Lang Publisher.

2.2 Vingt ans de recherche sur les perturbations de formants

Les perturbations de formants ont été introduites pour la première fois par Houde & Jordan (1998). Vingt ans plus tard, nous proposons un tour d'horizon de l'impact des travaux de Houde et Jordan au sein de la recherche sur la parole, et au-delà. Ces travaux sont cités notamment dans les domaines du mouvement du bras, ainsi que dans des études chez les animaux.

Parmi les 584 références citant Houde & Jordan (1998), extraites comme point de départ de notre analyse, 77 articles ou actes de conférence étaient consacrés à l'étude des réponses à des perturbations de formants. On y distingue la notion de compensation, ou changement de prononciation en réponse à des perturbations de formants ponctuelles et imprévisibles pour le

locuteur, et la notion d'adaptation auditorimotrice, qui a lieu en réponse à une perturbation auditive maintenue sur un certain nombre de répétitions (souvent plusieurs dizaines).

Après une synthèse des outils utilisés pour générer ces perturbations, une revue thématique détaillée est proposée, montrant les apports du paradigme de perturbations de formants pour la recherche sur la parole. Les perturbations de formants ont permis en particulier d'explorer les relations entre perception et production de la parole. Certaines études ont décortiqué le rôle du retour auditif dans la production de parole, ce qui a ensuite alimenté les modèles de production de la parole. L'altération conjointe des retours auditifs et somatosensoriels ont permis de mieux comprendre l'intégration de ces deux modalités sensorielles en production de parole. Associé à des tests et/ou des entraînements perceptifs, le paradigme de perturbation de formants a révélé des liens étroits entre contrôle de la parole et catégories phonologiques à travers la perception catégorielle. Les relations entre production de parole et unités phonologiques ont aussi été explorées avec l'étude de la généralisation ou transfert d'apprentissage auditorimoteur. Des études en imagerie cérébrale ont identifié les corrélats neuronaux de l'intégration sensorimotrice et du contrôle de la production de parole. L'altération de ce dernier observée dans certaines pathologies a pu être mieux comprise. La compréhension du développement de la production de parole a bénéficié aussi du paradigme. Enfin, certaines études ont mis en évidence l'influence de facteurs cognitifs et contextuels. Ces deux derniers points soulèvent des perspectives passionnantes pour les recherches en perturbations de formants, dont certaines sont suggérées en fin de chapitre.

Note : Le lecteur est invité à porter une attention particulière à la section concernant le transfert car elle est importante pour la suite de la thèse (section Research topics tackled with formant perturbations, sous-section intitulée Transfer/Specificity and speech units). En effet, c'est le transfert d'apprentissage auditorimoteur qui nous permet de questionner la nature des unités en production de parole. La sous-section précédant celle sur le transfert, consacrée aux catégories perceptuelles et phonologiques (Perceptual and phonological categories) est aussi particulièrement pertinente pour la suite de cette thèse, en particulier vis-à-vis du chapitre 4.

Tiphaine Caudrelier, Amélie Rochet-Capellan

Changes in speech production in response to formant perturbations: An overview of two decades of research

Abstract: One way to investigate speech motor learning is to create artificial adaptation situations by perturbing speakers' auditory feedback in real time. Formant perturbations were introduced by Houde and Jordan (1998), providing the first evidence that speakers adapt their pronunciation to compensate for these perturbations. Twenty years later, this chapter provides an overview of the general impact of Houde and Jordan's work in speech research and beyond, as well as a more detailed review of studies that involve formant perturbations. The impact of Houde and Jordan's work appears to be cross-disciplinary. Although mainly related to speech production and perception, it has also been cited in the limb movement and even animal research, mainly as evidence of adaptive sensorimotor control. Formant perturbations research has expanded rapidly since 2006, spreading across the world and many research teams. We identified 77 experimental studies focused on formant perturbations which we then analyzed with regard to technical and theoretical issues. This analysis showed that various apparatuses and procedures were used to address important topics of speech research. A primary interest has been in feedback and feedforward control mechanisms in speech. These mechanisms were addressed in different populations, including adults and children with typical vs. atypical development, with behavioral or neurophysiological approaches, or both. Some formant perturbations studies more specifically focused on the integration of auditory and somatosensory feedback in speech production, while others explored the interaction between speech production and perception of phonemic contrasts. Some research questioned the processes and the nature of speech representations by investigating generalization of adaptation to formant perturbations. Finally, a few studies were interested in the effect of extraneous variables such as surface effects or speakers' general cognitive abilities. Altogether, these studies provide insights into speech motor control in general and into the understanding of sensorimotor interactions in particular. The field has developed recently and may still expand in the future, as it allows us to address fundamental topics in speech research such as perception-production links or abstract vs. exemplar representations. Future

research with formant perturbations may also further connect sensorimotor adaptation to linguistic and cognitive factors and in particular to working and long-term memory.

Keywords: perturbation, real-time auditory feedback, formants, speech units, learning

1. Introduction

As an “extraordinary feat of motor control” (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984, p. 812), speech production is a challenging research topic, highly influenced by movement sciences (Grimme, Fuchs, Perrier, & Schöner, 2011; Maas et al., 2008). Speech motor control indeed shares numerous features with other sensorimotor systems and in particular with limb motor control. Among these features, sensorimotor adaptability of speech is of particular interest to speech science as the basis of speech rehabilitation (Maas et al., 2008), and since it is ubiquitous in daily life. Common examples include, among others, changes in the way we speak according to our interlocutor or to the surroundings, such as speaking louder when talking with someone with a hearing impairment or in a noisy environment (Garnier, Henrich, & Dubois, 2010); or spontaneously imitating our interlocutor’s speech sounds (Pardo, 2006). Speech motor control also adapts throughout the lifespan to natural or accidental alterations of our sensory systems or vocal tract geometry, temporarily or more permanently (Jones & Munhall, 2003; Lane et al., 2007). These adaptations allow maintenance of some level of intelligibility despite vocal tract growth, hearing loss, orofacial surgery, or when wearing a dental apparatus, losing teeth, speaking while eating etc. Being essential to speech production, sensorimotor adaptation of speech is the topic of numerous studies. For the purpose of this chapter, we will focus on studies that involved specific perturbation of formants. Formants are frequencies corresponding to peaks of acoustical energy, the relative values of which characterize vowels. Research in this field, and especially Houde and Jordan’s work, was inspired by the study of visuomotor adaptation in the limb movement literature (Houde & Jordan, 1998).

Pioneering work on adaptation of different visuomotor activities appeared at the end of the 19th century (Held, 1965; Stratton, 1897). This

work introduced a now common approach to assessing visuomotor adaptation that consists of investigating changes in movement in response to a systematic distortion of visual feedback, such as prism adaptation. As an illustration, Stratton (1897) reported his own and extreme everyday life experience while wearing an apparatus for eight days that reversed the retinal image upside down and left to right. On the first day, “the entire scene appeared upside down”. He felt nauseous. His movements were “laborious”, “embarrassed”, “inappropriate” (p. 344), required a lot of attention and were “extremely fatiguing” (p. 344). By the start of the third day things were much better, with no sign of “nervous distress” (p. 349). At the end of the fourth day, he “preferred to keep the glasses on rather than sit blindfolded” (p. 351/352). When the apparatus was removed on day eight, it took him some time to go back to normal feelings and motions.

Later work on visuomotor adaptation focused on more specific activities, less dramatic and more local and short-term changes, with a focus on reaching movements performed with rotations of the visual field. In this context, it has been repetitively demonstrated that when movements are achieved while the visual field is shifted by a specific angle (α), participants first miss the target by the same angle α . However, with repetition, they progressively learn to adapt their movements to the new feedback and reach the target accurately again. When they return to normal vision, after-effects and transfer effects are observed: participants miss the training target (*after-effects*) and/or a new target (*transfer*) by an angle more or less close to $-\alpha$. These effects vary as a function of the angular distance between the training and the testing targets (Krakauer, Pine, Ghilardi, & Ghez, 2000; Shadmehr & Mussa-Ivaldi, 1994). Sensorimotor adaptation has been attributed early on to feedforward control (i.e. predictive control based on learnt sensorimotor mappings) in contrast to forward closed-loop control (i.e. online processing of sensory inputs), visible in correction to unexpected perturbations (Golfinopoulos, Tourville, & Guenther, 2010; Houde & Chang, 2015). These notions are defined later in this chapter.

Twenty years ago, Houde and Jordan (1998) introduced an analogous procedure of visuomotor rotation adaptation to question feedforward control in speech, which used real-time alterations of formant frequencies in vowels. By altering the frequencies of the first and/or second formants (F1 and F2 respectively) it is possible to make a vowel sound like another

vowel. For example, by decreasing F1 and increasing F2, the vowel / ϵ / would sound closer to the vowel / i /, as illustrated in Figure 1. This alteration *displaces* the auditory feedback, in the same way as prism vision displaces the visual position of the target. For example, the speaker says “head”, speaking into a microphone and wearing headphones (Figure 1.A). The signal is processed in real time so that F1 and F2 formants are moved towards “hid” (Figure 1.B), and played-back into the headphones. The consequence for the speaker is a discrepancy between the auditory target expected from the planned movements (“head”) and the auditory target they actually got (~“hid”). In other words, similar to visuomotor adaptation, the speaker first *misses* the auditory target (Figure 1.C, “Training start”). With practice – repetition of shifted utterance(s) with the same perturbation – the speaker adapts to the perturbation (Figure 1.C, “Training end”): To *reach* the auditory target “head” again in the presence of the perturbation, they produce formants in the opposite direction to the perturbation. In our example, this corresponds to the production of an utterance closer to “had”. When the feedback is returned to normal or masked with a noise, for the same vs. different utterance(s) than the training one(s), after-effects vs. transfer effects are observed (Figure 1.C, column “After-effect” and “Transfer”). This suggests that the compensation is not only an online feedback control change but also affects auditory-motor mappings supporting feedforward control, in a more or less utterance or segment-specific way. The procedure was later adapted to address feedback control by investigating online compensation to unexpected perturbations (Purcell & Munhall, 2006b).

Adaptation to formant perturbations has been investigated per se, or used as a paradigm to address more general issues in speech science. The current chapter reviews research in formant perturbations by analyzing Houde and Jordan’s seminal study (Houde & Jordan, 1998, 2002) and the scientific literature that has referred to it. Using this approach (detailed in the first section of the chapter) we can see the cross-disciplinary impact of Houde and Jordan’s work and in particular, identify the main topics of the scientific literature that have cited this work (reported in the second part of the chapter). Among the collected papers, only a subsection corresponded to empirical studies involving formant perturbations. Based on the analysis of these studies, including review of their reference lists, the latter parts

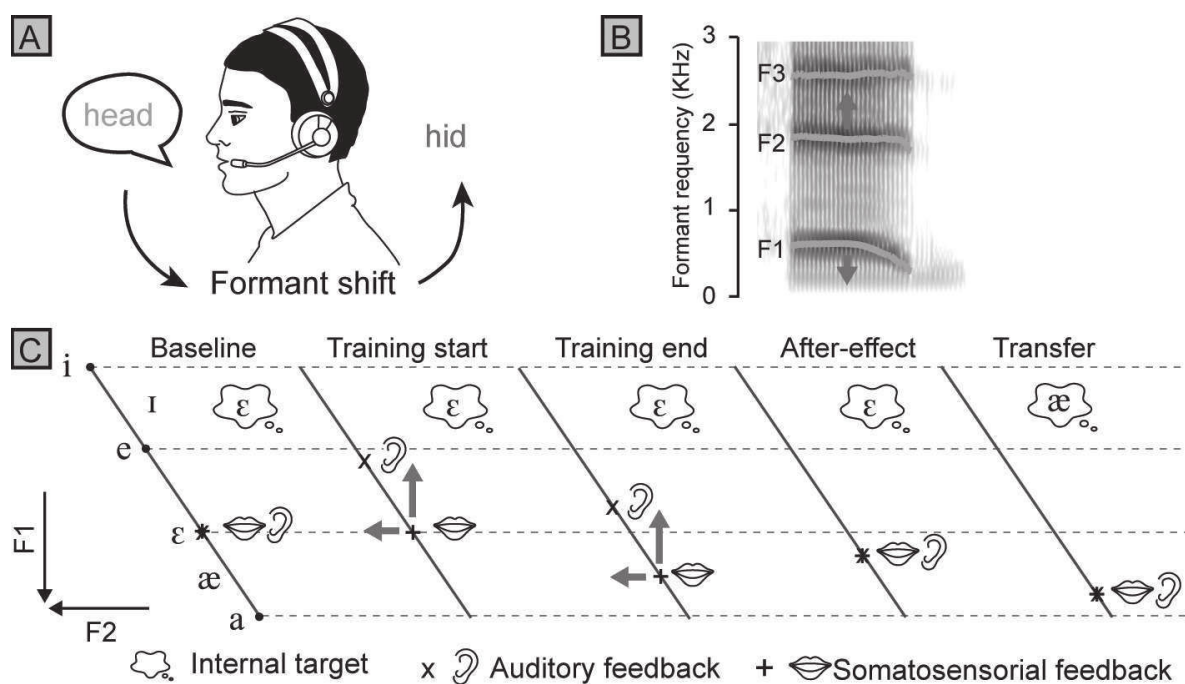


Figure 1: The auditory prism adaptation. (A) The speaker speaks into a microphone; his feedback is altered such as when he produces “head” he is hearing a signal closer to “hid”; (B) To do so, F1 and F2 are changed in real time; (C) Before the introduction of the perturbation (Baseline) the auditory feedback is consistent with the target. The first exposure to the perturbation (Training start) induces a discrepancy (or an error) between the auditory feedback and the planned target. With repetitive exposure to the perturbation, the talker changes his production to compensate for the perturbation (Training end). When the perturbation is removed after-effects and/or transfer effects are observed.

of the chapter provide: (1) a description of the main apparatuses and paradigms used in formant perturbations studies; (2) an overview of the research topics addressed using these perturbations and the main reported results; and (3) some perspectives for future research.

2. Paper collection and analysis

As we were interested in the impact of Houde and Jordan’s work and also wanted to provide an analytical review of formant perturbations studies, we first analyzed the published work that referred to Houde and Jordan (1998 and/or 2002) from 1999 to 2018 (last update on July 6th 2018). This was performed using the “Cited by” function in Google Scholar. We choose this approach rather than keyword research, as we wanted to collect various sorts of publication, and because it appeared to be the

Table 1: Number of references in each category of the first level of selection (see text for details)

Formant shift		No formant shift		Not in English	Error ref.	Total
Rejected	Kept	Rejected	Kept			
26	72 (+ 2, Houde & Jordan 1998 and 2002)	140	287	35	22	584

most systematic way to collect publications in the field. To compensate for potential errors and omissions by Google Scholar, the results were then analyzed very closely.

An analysis by year of Google Scholar output resulted in a total of 584 references (including the two papers by Houde and Jordan, see Table 1). As a first step, we excluded documents that were not written in English or that corresponded to reference errors (57 in total, see Table 1). Among the 527 remaining references, we distinguished between those without vs. with an empirical study that included formant perturbations. In the former category (n=427, without formant perturbation), we kept only journal papers for a thematic analysis of Houde and Jordan's broad impact (n=287). In the latter category (n=100, with formant perturbations), we first kept all the documents except PhD or Master theses, posters or abstracts to conferences (74 references kept, 26 rejected). Note that there were 11 PhD theses; most of them were associated with journal publications. For consistency in criteria, we did not include Frank (2011)'s PhD thesis, even though it is often cited by studies investigating linguistic effects on formants adaptation and its results were never published in peer-reviewed papers.

Three more papers were added that included formant perturbations. One paper that did not cite Houde and Jordan was found in the reference list of the selected papers (Niziolek & Guenther, 2013); and two papers in course of publication at the time of writing that we were aware of (Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2018; Klein, Brunner, & Hoole, in this book). The general characteristics of the documents including formants perturbations are described in Table 2. Technical

Table 2: Number of papers considered for the analysis of formant perturbations according to source and type. Houde & Jordan (1998, 2002) are included.

	Journal papers	Proceedings papers	Reports/ chapters	Total
Google Scholar	55	17	2	74
Other sources	1	1	1	3

papers as well as papers investigating compensation to unexpected formant perturbations were included.

The full list of analyzed papers related to formant perturbation is available in Table 4 at the end of this chapter, with their main related research topic indicated. As the paper collection is based mainly on the “cited by” function of Google Scholar some papers may be missing despite our careful attention. However, we believe our analysis provides an accurate picture of the field at the time it was run.

3. Overall impact of Houde and Jordan’s seminal work

The overall impact of Houde and Jordan (1998, 2002) is illustrated in Figure 2. We distinguished seven broad categories of research: (1) formant perturbations studies (n=77); (2) studies that investigated speech compensation and/or adaptation to other auditory perturbations or equivalent situations (n=91) or (3) to an alteration of the vocal tract (n=16); (4) empirical or theoretical papers on speech production (n=61) or (5) on speech perception (n=46); (6) studies involving non-speech actions (n=25); and (7) experimental or theoretical papers involving animals (n=43). Five papers were not considered, as they were difficult to classify in these categories. We first analyzed the journal papers that did not empirically test formant perturbations. As described above, this involved 286 articles. Broad research topics were identified mainly from abstract reading. A subset of papers was selected and read in more detail to illustrate the different topics. The articles on formant perturbations will be reviewed in detail in the next sections. We will now briefly overview the research topics in the six other categories. References in the following section are illustrative.

3.1. Compensation/adaptation of speech production to various auditory perturbations

Speech compensation and adaptation were investigated prior to the development of formant perturbation studies and used various methods. These methods continued to be used in some of the later work that cited Houde and Jordan. About half of the papers in this first category investigated speech modifications in reaction to either an unexpected or a predictable modification of F0 in different populations and conditions. A number of papers in this topic were published by Jones et al. (Jones & Munhall, 2000); Larson et al. (Burnett & Larson, 2002); or Hanjun et al. (Li et al., 2016). The other half of the studies investigated speech modifications in reaction to other types of auditory perturbations such as delayed auditory feedback (Chon, Kraft, Zhang, Loucks, & Ambrose, 2013); changes in intensity or noise level (Maas, Mailend, & Guenther, 2015); hearing loss (Palethorpe, Watson, & Barker, 2003); real or simulated use of cochlear implants (Casserly, 2015; Lane et al., 2007); or replacement of the auditory feedback by a stranger's voice (Hubl et al., 2014). Other work modified consonant features such as frication (Shiller, Sato, Gracco, & Baum, 2009) or voicing (Mitsuya, MacDonald, & Munhall, 2014). Self-regulation in adaptation to formant perturbations was also linked with interpersonal auditory-motor regularizations in speech such as phonetic convergence (Pardo, 2006).

3.2. Compensation/adaptation of speech production to perturbations of the vocal tract dynamics or geometry

Research on compensation and adaptation to perturbations affecting the somatosensory feedback is another field closely connected to adaptation to formant perturbations. Houde and Jordan's work was thus cited by studies involving an alteration of the vocal tract geometry or dynamics. This includes dental prostheses (Jones & Munhall, 2003); lip tubes in children and adults (Ménard, Perrier, & Aubin, 2016); false palates (Thibeault, Ménard, Baum, Richard, & McFarland, 2011); mechanical forces applied to the jaw with a robot (Tremblay, Shiller, & Ostry, 2003); or more permanent changes such as those induced by oropharyngeal cancer treatments (de Bruijn et al. 2012).

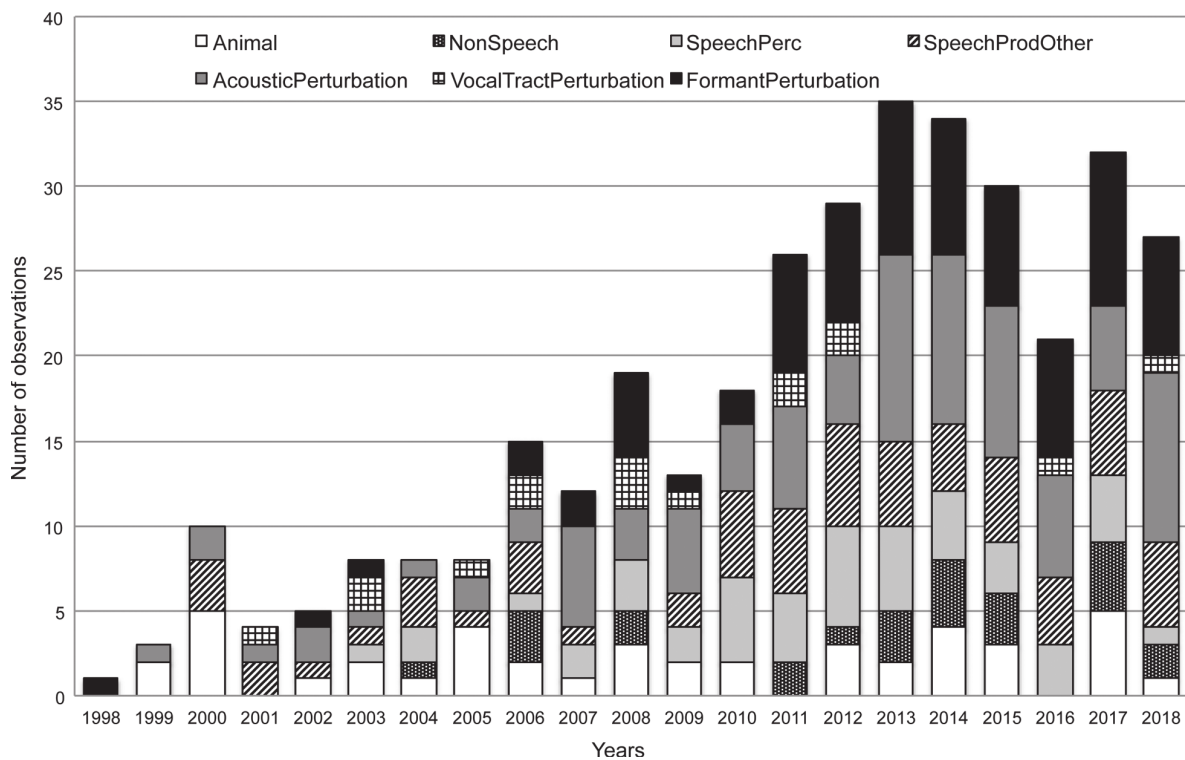


Figure 2: Overall impact: number of analyzed papers by year and categories.

3.3. Empirical or theoretical papers on speech production

Houde and Jordan's work is cited by empirical and theoretical research on speech production. For example, adaptation to formant perturbations is mentioned by studies providing further evidence of the role of auditory feedback in speech motor control, such as work linking auditory acuity to the production of speech contrasts (Perkell et al., 2004); auditory perceptual learning with improvement in production (Shiller, Rvachew, & Brosseau-Lapr e, 2010); comparing overt and covert speech (Brumberg et al., 2016) or analyzing the neurophysiological activities of the auditory cortex during speech production (Curio, Neuloh, Numminen, Jousm aki, & Hari, 2000). Adaptation to formant perturbations provides support for neurocomputational models of speech production such as the Directions Into Velocity of Articulators model (DIVA, Golfopoulos et al., 2010) or the State Feedback Control model (SFC, Houde & Chang, 2015), both models assuming a feedback and a feedforward control mechanism. Further information about these control mechanisms will be provided in the section describing formant perturbation studies related to this topic.

3.4. Empirical or theoretical papers on speech perception

Adaptation to formant perturbations is also taken as evidence of sensorimotor integration in speech. As such, it is relevant for papers probing or discussing the role of the motor system in speech perception (Sato, Troille, Ménard, Cathiard, & Gracco, 2013) or in theoretical papers related to the dual-stream model of language processing. Basically, this model proposes a cortical ventral stream that maps speech sounds to concepts, and a dorsal stream for auditory-motor mapping. Adaptation to formant perturbations is then cited as an evidence that a dorsal auditory-motor integration path is still functional in adulthood (Hickok & Poeppel, 2004).

3.5. Non-speech movement studies

Various non-speech studies cited Houde and Jordan's work to illustrate sensorimotor adaptation in humans. These studies focused on activities involving auditory feedback such as piano playing (Pfordresher & Palmer, 2006); or the learning of artificial auditory-arm movement maps (van Vugt & Ostry, 2018). Some papers were also interested in other kinds of sensorimotor adaptations such as swallowing (Wong, Domangue, Fels, & Ludlow, 2017), or visuomotor adaptation of limb movements (Wei et al., 2014). Note that as formant perturbations studies were inspired by visuomotor adaptation, they often referred to limb movement literature. The converse seems not necessarily true as our research suggests that few works on limb adaptation have cited Houde and Jordan's work. This result should be taken cautiously as limb movement research could cite other studies using formant perturbations to illustrate the adaptability of speech motor control, and we only collected papers that reference Houde and Jordan using "cited by" functionality of Google Scholar.

3.6. Animal studies

Finally, animal studies have early, and regularly, cited Houde and Jordan's work (Figure 2), with a main focus on the role of auditory feedback in action control. Over half of these papers were dedicated to birdsong and published by Brainard et al. and/or Doupe et al. and/or Sober et al. Many of these papers include studies of birdsong production or learning using auditory perturbations with behavioral and/or neurophysiologic recordings, as

well as interspecies comparative reviews about the processing of auditory feedback of self-produced sounds (Brainard & Doupe, 2000; Doupe & Kuhl, 1999; Sober & Brainard, 2009). Analogous works were done in bats (Smotherman, Zhang, & Metzner, 2003) and primates (Eliades & Miller, 2017).

To summarize, this non-exhaustive analysis of the overall impact of Houde and Jordan's seminal work suggests that it is (as expected) cited by papers investigating speech compensation and adaptation to other types of sensory perturbations. Most of the scientific questions in this first set of papers overlap with the research topics we will review based on the more detailed analysis of formant perturbations studies in the related section of this chapter. In a broad context, adaptation to formant perturbations is often interpreted as evidence for sensorimotor integration and sensorimotor plasticity in speech production and perception. It is cited to illustrate auditory feedback and feedforward control mechanisms in speech production, as explained below, and taken as an example of such mechanisms (and their plasticity) in studies investigating animal vocalizations, singing, music playing, but also inter-personal convergence or coordination of movements.

Note that more research topics related to formant perturbation studies may be found by including "2nd order" connections to Houde and Jordan's work (i.e. references that cite any of the studies on formant perturbations).

4. Methods in formant perturbation studies

In this section, we provide an overview of the apparatuses used to apply real-time formant perturbation and a description of the main procedures identified in the collected papers.

4.1. Real-time formant perturbation

The systems used to shift formants in the collected papers are summarized in Table 3. Paper details can be found in Table 4. With regards to formant perturbation, it is important to emphasize that in order to preserve the best quality of self-perception, the real-time modification of formants in speakers' auditory feedback should meet some requirements, specifically:

- (1) The signal should be processed and played back fast enough for the speaker not to perceive any delay (less than 30ms, see Yates, 1963). Specific digital signal processing boards (DSP), including systems from the music industry were used, especially in earlier work. Nowadays, this can be achieved at a software level, on a PC with appropriate sound card and software to analyze and change formants. For the same code, the achieved delay can vary depending on the operating system and hardware.
- (2) The parameters of the signal processor should be adapted to the speaker and/or to the vowel. This parameterization improves the formant detection and the reliability of the perturbation.
- (3) Perception of unperturbed feedback (bone conduction and air conduction outside the headphones) should be reduced as much as possible. Different approaches were used to achieve this aim, such as:
 - Using whispered speech (Houde & Jordan, 1998, 2002) although subsequent studies were run with normal speech;
 - Using closed headphones or insert earphones to reduce the perception of the air-conducted signal. The occlusion effect of the headphones on adaptation was recently investigated with no significant difference in the magnitude of F1 adaptation between the use of the closed Sennheiser “HD 265” and the insert Etymotic Research ER2 (Mitsuya & Purcell, 2016);
 - Increasing the level of the feedback in the headphones, up to 87dB SPL (Villacorta et al., 2007);
 - And/or using a masking noise mixed with the played back signal to mask bone-conducted speech.
- (4) The shifted vowel should have clearly distinguishable F1 and/or F2 values, and the shift should be consistent with these values. For this reason, the vowel / ϵ / is chosen in most of the studies as shifting more extreme front or back vowels could be limited by overlap in F1–F2 or F0–F1 frequencies (Mitsuya, MacDonald, Munhall, & Purcell, 2015), and this vowel allows upward and downward perturbations.

Different research groups have developed their own formant perturbation systems (Table 3) with four main categories: (1) The two systems developed by Houde described with more details in Houde’s PhD (Houde, 1997) for whispered speech (1.a), and then in Katseff, Houde, & Johnson

(2012) for voiced speech (1.b); (2) The system developed and used by Munhall, Purcell and collaborators that used a specific hardware; (3) The system used by Perkell and Guenther’s teams that first included specific hardware (Villacorta et al., 2007) and was then adapted as a free software for Matlab. It supports various auditory perturbations, including changes in F1 and/or F2, but also more complex ones such as formant trajectory perturbations (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Tourville, Cai, & Guenther, 2013). The last version is called “Audapter” and can be download on github.com (https://github.com/shanqing-cai/audapter_matlab, this link was retrieved July, 6, 2018); (4) The last system was developed in parallel by three teams: Max et al., Ostry et al., and Shiller et al. It uses a device from the music industry (VoiceOne, TC Helicon) that by default allows shifting of all the formants while preserving F0. This system was used as a way to alter all formants in the same direction (Max & Maffett, 2015) or, with supplementary signal processing steps, including filtering and mixing, as a way to perturb F1 only (Rochet-Capellan & Ostry, 2011). A few papers were dedicated to the presentation and first

Table 3: Main signal processing systems used in the literature to perturb formants in real time (references indicate the publication describing the system) and number of papers using the system.

	System 1	System 2	System 3	System 4	
References	Houde (1997); Katseff et al. (2012)	<i>Purcell & Munhall (2006ab)</i>	<i>Villacorta et al. (2007); Cai et al. (2008); Tourville & al. (2013)</i>	Feng et al. (2011); Rochet-Capellan & Ostry (2011); Shum et al. (2011)	Others
Signal processing	1.a. Whispered speech: Analysis-synthesis process, DSP- 96 board, Ariel, Inc. 1.b.Voiced speech: “Feedback Alteration Device” – Sinewave synthesis	National Instruments PXI-8176 embedded controller	Texas Instruments C6701 Evaluation Module DSP board then C-extension Mex for Matlab, opened access – Audapter	Electronic speech processor from music industry VoiceOne; TC Helicon + filters	Other software or hardware solutions –
Number of papers	10	23	2 then 20	19	3

evaluation of these different perturbation systems. This was the case with Cai et al. (2008) and Tourville et al. (2013) and with the preliminary work by Shih, Suemitsu, & Akagi (2011). Two papers also presented a method to perturb formants in populations in which speech acoustics have deteriorated, by coupling articulatory synthesis with Audapter (Berry, North, & Johnson, 2014; Berry, North, Meyers, & Johnson, 2013).

As displayed in Table 4, most of the studies involved native speakers of English, mainly from North America. Other languages were investigated in a few comparative studies or in relation to other research questions as described in the next section. Potential generalization of these findings to other languages and populations should therefore be taken with caution.

4.2. Main procedures in formant perturbation studies and related concepts

The main procedures identified in the collected papers about formant perturbations are summarized in Figure 3. These procedures will be referred to in relation to the research topics detailed in the next section. Two main approaches can be distinguished:

- (1) Unexpected formant perturbation during the production of prolonged utterances: This first approach was used in only a few of the collected papers ($n=11$, ~14 % of the papers with formant perturbations, see Table 4). The perturbation is only applied to a small proportion of utterances so that talkers cannot anticipate the perturbation. Moreover, the utterances are produced with long vowel duration (steady-state vowels) so that corrective answers result from online processing of the auditory feedback (cf. Figure 3, procedure P4). This correction is called *compensation*.
- (2) Systematic and constant perturbation over a number of utterances: This second approach was used in the majority of the papers ($n=66$, ~86 %, Table 4). The basic procedure is represented in Figure 3, procedure P1. It involves the production of utterances with “natural” duration, in general. After a baseline with unaltered auditory feedback, the perturbation is introduced either gradually or abruptly, and then systematically applied at a constant level. Depending on the research group, changes in formant production

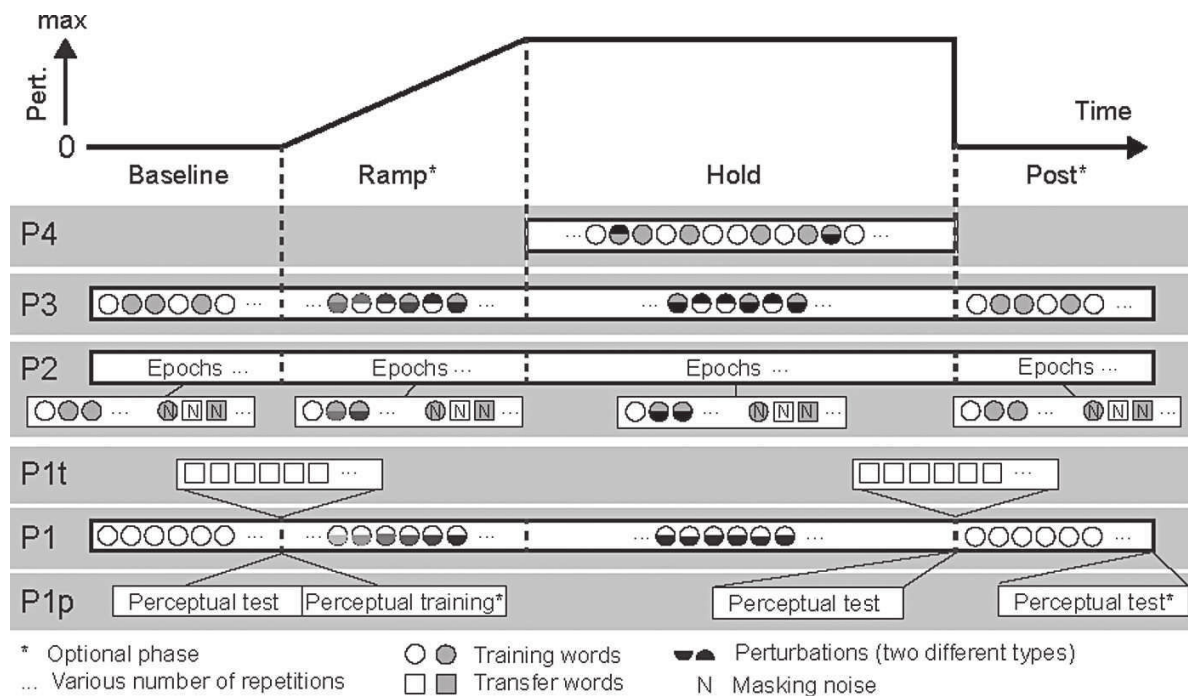


Figure 3: Overview of procedures used in formant perturbations studies. Duration of experimental phases and perturbations were variable across studies. P1 is the basic procedure to study auditory-motor adaptation, used in Munhall et al.'s studies. It was adapted to investigate the transfer of adaptation (P1t) (MacDonald, Pile, Dajani, & Munhall, 2008; Rochet-Capellan, Richer, & Ostry, 2012) and the effect of auditory motor adaptation on perception (P1p) (Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014) or the effect of perceptual training on sensorimotor adaptation (Lametti, Krol, Shiller, & Ostry, 2014). P2 is the procedure used in Houde & Jordan (1998) and then by Perkell et al. (Villacorta, Perkell, & Guenther, 2007). It is structured in epochs with training words produced with feedback followed by training words and generalization words produced with a masking noise. P3 is the multiple perturbation procedure developed in Rochet-Capellan & Ostry (2011), during which words are produced in random order with specific perturbation associated with each word. P4 is the compensation procedure to unpredictable perturbations. In this last case, long steady-state vowels are produced and the perturbation is introduced randomly for a small proportion of utterances to assess online correction (Purcell & Munhall, 2006b). Grey scale gradient in the ramp phase represents the progressive introduction of the shift.

at the end of the training phase are referred to as *compensation* (cf. Houde & Jordan, 1998; Purcell & Munhall, 2006b) or *adaptation* (cf. Rochet-Capellan, Richer & Ostry, 2012, Martin et al., 2018), and residual changes when the feedback is returned to normal after training are referred to as *adaptation* or *after-effect*, respectively.

This procedure was also used to assess *generalization* (or *transfer*) of adaptation to untrained utterances, either in the course of the training phase (Figure 3, procedure P2) or after the training (Figure 3, procedure P1t), as presented in the next section.

Hereafter, *adaptation* will refer to changes observed at the end of the training phase in response to a systematic perturbation. *Compensation* will mainly refer to changes in response to unpredictable perturbations but will also be used to qualify the direction of adaptive responses (by contrast with following responses that go in the same direction as the perturbation).

5. Research topics tackled with formant perturbations

In this section, we provide a thematic review of the collected papers that included an empirical study of formant perturbation. As much as possible, we chose to associate each paper with a main topic but obviously a paper could be related to more than one topic. Table 4 provides a list of all the cited references and their main associated research topics.

5.1. Properties of feedback and feedforward control

Many studies involving formant perturbations are related to the role of auditory feedback in speech motor control and distinguish between feedback and feedforward control mechanisms. Feedback control is a closed-loop system that involves the sensory consequences of the current motion. It is regarded as too slow to account for rapid control and rapid adjustments observed in fast coordinated actions. Rapidity and adaptability of motion were identified early on as evidence of a feedforward control mechanism by researchers in visuomotor adaptation. The core idea is that the brain makes predictions of the sensory consequences of its actions based on an efference copy of the motor command (Houde & Jordan, 2002). These predictions involve mappings between motor and sensory representations also called *internal models* (Purcell & Munhall, 2006a) or sensorimotor memories (see Perrier, 2012, for a discussion of the nature of internal models in speech). The DIVA (Golfinopoulos et al., 2010) or the SFC (Houde & Chang, 2015) neurocomputational models of speech production assume the existence of both feedback and feedforward control networks that involve auditory and somatosensory systems. When

the prediction based on *internal models* does not match the actual sensory input, the internal representations are changed to reduce this prediction “error” so that future movements performed in similar conditions will be accurate. This mechanism is claimed to underlie sensorimotor adaptation.

In this context, a first subset of studies with formant perturbations was designed to “Investigate the nature, level of details, and use of internal models in speech production” (Max, Wallace, & Vincent, 2003, p. 1053) and to “begin to parameterize the formant feedback system” (MacDonald, Goldberg, & Munhall, 2010 p. 1060). The main contribution of these studies is to describe the role of auditory feedback in the control of formant production, and the adaptability of this control. In these papers, adaptability is mainly explained or taken as an evidence for feedforward internal models.

To address the properties of adaptation to formant perturbations, Houde and Jordan (2002) analyzed in more detail the adaptation phenomenon introduced in Houde and Jordan (1998). The results highlight some properties of feedback and feedforward control that were subsequently discussed and investigated in later work, involving various types of formant perturbations and procedures.

The first observation of Houde and Jordan was that the changes in F1 and F2 production in talkers’ speech were compensatory responses, in the opposite direction to the perturbation. This result has been reproduced consistently in later work when between-speaker data are aggregated. Individual data suggests that some speakers follow the shift, however. For example, in a meta-analysis of their own studies of adaptation to formant perturbations, MacDonald et al. (2011) found that 26 out of 116 female speakers followed F1 or F2 shifts when their production of “head” was perturbed toward “had”. A possible explanation is that non-adapted speakers may not be able to dissociate their own production from the auditory feedback (Vaughn & Nasir, 2015). Following the formant shift rather than compensating for it was actually the most frequent behaviour observed in a preliminary study investigating compensation in Japanese speakers to unexpected perturbations of F1, F2 and F3 (Shih et al., 2011). Aside from this study, all other published work on formant perturbations observed significant compensatory adaptation in acoustic analyses, whereas preliminary analyses of articulatory correlates of adaptation are

Table 4: List of all the studies related to formant perturbation included in the present review. The first column provides the reference of the article. The 2nd column gives the language of participants (Du: Dutch, En: English, Fr: French, Ge: German, Ja: Japanese, Ko: Korean, Ma: Mandarin, Ru: Russian, Sp: Spanish). Column 3 is related to the perturbation systems, which are described in Table 3 (briefly, 1.a: Houde & Jordan (1998), 1.b. Katseff et al. (2012); 2: Purcell & Munhall, (2006a); 3: Audapter and its previous versions; 4: VoiceOne, TC Helicon, 5: Others) and column 4 indicates whether an article is mainly dedicated to the description of a perturbation system. Each study has been classified into either compensation (to unpredictable perturbations, column 5) or adaptation (to sustained perturbations). Columns 7 to 14 show whether the article is related to each of the main research topics presented in the present review. A cross indicates that the article is cited in the corresponding subsection, while a (X) indicates it is not although it is related to the topic.

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Alsius, Mitsuya, Latif, & Munhall, 2017	En	2			X		(X)						X
Berry, Jaeger, Wiedenhoeft, Bernal, & Johnson, 2014	En	3			X	X			X				
Berry, North, & Johnson, 2014	En	3	X										
Berry, North, Meyers, & Johnson, 2013	En	3	X										
Bourguignon, Baum, & Shiller, 2014	En	4			X			X					
Bourguignon, Baum, & Shiller, 2015	En	4			X			X					
Bourguignon, Baum, & Shiller, 2016	En	4			X			X					
Cai, Beal, Ghosh, Tiede, Guenther, & Perkell, 2012	En	3			X		(X)		X				

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Cai, Boucek, Ghosh, Guenther, & Perkell, 2008	Ma	3	X		X								
Cai, Ghosh, Guenther, & Perkell, 2010	Ma	3			X	X		X					
Cai, Ghosh, Guenther, & Perkell, 2011	En	3		X		X							
Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2016	Fr	3			X			(X)	X				
Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2018	Fr	3			X				(X)				X
Caudrelier, Schwartz, Perrier, Gerber, & Rochet-Capellan, 2018	Fr	3			X			(X)	X				
Daliri, Wieland, Cai, Guenther, & Chang, 2018	En	3			X		(X)		X				
Lametti, Krol, Shiller, & Ostry, 2014	En	4			X			X					
Lametti, Nasir, & Ostry, 2012	En	4			X		X						
Lametti, Smith, Freidin, & Watkins, 2018	En	4			X						X		
Demopoulos et al., 2018	En	1b		X	X					X		(X)	
Deroche, Nguyen, & Gracco, 2017	En	4			X		(X)				X		
Dimov, Katseff, & Johnson, 2012	En	1b			X								X
Eckey & MacDonald, 2015	Ge	5		X			X						
Feng, Gracco, & Max, 2011	En	4			X		X						
Houde & Jordan, 1998	En	1a			X	X			X				

(continued on next page)

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Houde & Jordan, 2002	En	1a		X	X								
Ito, Coppola, & Ostry, 2016	En	4		X			(X)				X		
Katseff & Houde, 2008	En	1b		X			(X)						
Katseff, Houde, & Johnson, 2012	En	1b		X			X						
Klein, Eugen; Brunner, Jana; Hoole, Phil (sous press)	Ru	3		X					X				(X)
Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014	En	4		X				X					
MacDonald & Munhall, 2012	En	2		X			X						
MacDonald, Goldberg, & Munhall, 2010	En	2		X		X	X						
MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012	En	2		X								X	
MacDonald, Pile, Dajani, & Munhall, 2008	En	2		X					X				
MacDonald, Purcell, & Munhall, 2011	En	2	X			X							
Martin et al., 2018	Sp	1b		X			X						
Max & Maffett, 2015	En	4		X		X							
Max, Wallace, & Vincent, 2003	En	5		X		X							
Mitsuya & Purcell, 2016	En	2		X		X							
Mitsuya, MacDonald, Munhall, & Purcell, 2015	En	2		X			X						
Mitsuya, MacDonald, Purcell, & Munhall, 2011	En	2		X				X					
Mitsuya, Munhall, & Purcell, 2017	En	2				X							

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Mitsuya, Samson, Ménard, & Munhall, 2013	Fr	2			X			X					
Mollaei, Shiller, & Gracco, 2013	En	4			X					X			
Mollaei, Shiller, Baum, & Gracco, 2016	En	4		X			(X)			X			
Munhall, MacDonald, Byrne, & Johnsruide, 2009	En	2			X	X							(X)
Neufeld, Purcell, & Van Lieshout, 2013	Ko	2			X	X							
Niziolek & Guenther, 2013	En	3		X				X					
Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017	En	1b		X	X					X			
Pile, Dajani, Purcell, & Munhall, 2007	En	2			X				X				
Purcell & Munhall, 2006a	En	2			X	X							
Purcell & Munhall, 2006b	En	2		X		X							
Purcell & Munhall, 2008	En	2			X	X	X						
Reilly & Dougherty, 2013	En	3		X			X	(X)					
Reilly & Pettibone, 2017	En	3			X				X				
Rochet-Capellan & Ostry, 2011	En	4			X				X				
Rochet-Capellan, Richer, & Ostry, 2012	En	4			X				X				
Sato & Shiller, 2018	Fr	3			X		(X)				X		X
Schuerman, Nagarajan, & Houde, 2015	En	1b			X			X					
Schuerman, Nagarajan, McQueen, & Houde, 2017	En	1b			X			X					

(continued on next page)

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Schuerman, Meyer, & McQueen, 2017	Du	3			X			X			(X)		
Sengupta & Nasir, 2015	En	2			X						X		
Sengupta & Nasir, 2016	En	2			X						X		
Sengupta, Shah, Gore, Loucks, & Nasir, 2016	En	2			X					X	(X)		
Shih, Suemitsu, & Akagi, 2011	Ja	5		X		X							
Shiller & Rochon, 2014	En	4			X			X					(X)
Shiller, Lametti, & Ostry, 2013	En	4			X			X					
Shum, Shiller, Baum, & Gracco, 2011	En	4			X						X		
Terband & Van Brenk, 2015	Du	3			X								X
Terband, Van Brenk, & van Doornik-van der Zee, 2014	Du	3			X			(X)		X			(X)
Tourville, Cai, & Guenther, 2013		3	X										
Tourville, Reilly, & Guenther, 2008	En	3		X							X		
Trudeau-Fisette, Tiede, & Ménard, 2017	Fr	2			X		X				(X)		
van den Bunt, Groen, Ito, Francisco, Gracco, Pugh, & Verhoeven, 2017	Du	4			X			(X)		X			
Vaughn & Nasir, 2015	En	2			X	X							
Villacorta, Perkell, & Guenther, 2007	En	3			X	X	X		X				
Zheng, Vicente-Grabovetsky, MacDonald, Munhall, Cusack, & Johnsrude, 2013	En	2		X							X		

less clear. Max et al. (2003) analyzed acoustic changes to perturbation of all formants in the same direction in relation to jaw and tongue movement during adaptation. No consistent behaviour were observed in articulatory kinematics. Similar results were obtained in a pilot study in one Korean speaker with an F2 shift (Neufeld, Purcell, & Van Lieshout, 2013), while clearer tongue compensation movements were reported in speakers with blindness (Trudeau-Fisette, Tiede, & Ménard, 2017). On the other hand, while the majority of studies on adaptation to formant perturbations found significant compensatory responses, it was also shown that adaptation vanishes when perturbed feedback is delayed by more than 100ms (Max & Maffett, 2015), or is at least largely reduced (Mitsuya, Munhall, & Purcell, 2017).

Houde and Jordan also reported that maximal changes at the end of training did not fully compensate for the perturbation. This result was systematically reproduced in later studies. As an illustration, in Purcell & Munhall (2006a), the maximal adaptation to a 200Hz upward vs. downward shift of F1 compensated for about 30 % of the perturbation, regardless of the number of repetitions during the hold phase. This also suggests that adaptation is a fast process, in agreement with Max et al. (2003)'s observation that compensatory responses occurred after only a few repetitions. However, a F1 perturbation of at least 60Hz (80Hz on average across conditions) was required in Purcell & Munhall (2006a) to initiate the compensatory response. Similar thresholds were reported in later work, regardless of the delay in the auditory feedback (Mitsuya et al., 2017) and the occlusion of the headphones (Mitsuya & Purcell, 2016). Furthermore, MacDonald et al. (2010) highlighted a linear relationship between the magnitude of the perturbation and the magnitude of changes in speakers' utterances for perturbation magnitudes up to +200Hz in F1 and -250Hz in F2, compensating for 25 % of the perturbation in F1 and 30 % in F2. With larger perturbations, there was no improvement, and a decrease even appeared in response to perturbations larger than 300Hz in F1 and larger than 400Hz in F2. Similar limits were observed by Katseff and colleagues (Katseff & Houde, 2008; Katseff et al., 2012), as discussed in the next section. Comparable adaptations were reported in the meta-analysis provided by MacDonald et al. (2011), with an average of 26.5 % for F1 and 23.2 % for F2. Moreover, in this last analysis, changes in F1 in

speakers' production weakly correlated with changes in F2, suggesting a specific control of the two parameters and the existence of speaker-specific strategies. The magnitude of the response was also found to vary according to the vowel in *pet*, *bus* and *law* utterances in Max et al. (2003). Further work addressing this last point with regard to more specific research topics is presented in the next section.

Houde and Jordan also noticed that inter-speaker variability was not related to a speaker's awareness of the auditory shift. When interviewed after the study, talkers reported they were unaware of the perturbation or of any change in their production. By contrast, Purcell & Munhall (2006a) reported that 40 % of their participants indicated awareness of "some kind of change in the auditory feedback over the course of the experiment", with only 8 % noticing that the perturbation transformed the vowel into a different one. However, the magnitude of adaptation did not seem to be related to the responses in this interview. This difference to Houde and Jordan might be related to the abrupt suppression of the perturbation after training in Purcell & Munhall (2006a) (Procedure P1, Figure 3) that was probably perceived by the speakers, while Houde and Jordan assessed how adaptation was sustained using catch trials with masking noise (Procedure P2, Figure 3). Munhall, MacDonald, Byrne, & Johnsrude (2009) then confirmed that the awareness of the perturbation does not influence adaptive behavior, as discussed later in the "*Surface effects & speakers' characteristics*" subsection.

Another important result in Houde and Jordan was that changes for perturbed utterances were larger than changes for utterances produced with a masking noise. The authors discussed this result as evidence that "vowel production could be partly under immediate auditory feedback control" (Houde & Jordan, 2002, p. 307). By contrast, in their preliminary study of adaptation to a shift of all formants in the same direction, Max et al., (2003) argued that the modifications in talkers' production should be considered as adaptive responses rather than reactive changes, as they already occur at vowel onset, and have been observed for sustained vowels as well as vowels with shorter duration. The variability of changes in formants according to the vowel's parts were not systematically investigated in adaptation studies as most of the studies used a single steady-state value, often around the middle of the vowel. However, in their preliminary

work, Berry, Jaeger, Wiedenhoeft, Bernal, & Johnson (2014) suggested that this single value might not be the most appropriate, depending on consonant context and coarticulatory effects. Vaughn and Nasir (2015) also provided evidence that full trajectory analysis might better capture adaptation phenomena. The relationship between formant values in consecutive trials (as measured with one-lag cross correlation analyses), in the absence of any perturbation, may also be predictive of adaptation magnitude (Purcell & Munhall, 2006a). Altogether, these results suggest that changes observed over the course of adaptation to a perturbation result probably from a mix of feedback and feedforward control.

Houde and Jordan (2002) suggested investigating compensation to formant perturbations in steady-state vowels to determine the role of online feedback in formant control. Studies focusing on compensation to an unexpected formant perturbation in sustained vowels usually analyzed changes at different points of the vowel. For instance, in Purcell and Munhall (2006b) upward vs. downward perturbations of F1 were applied randomly in five utterances of “head” over 100 utterances of different CVC words. Results show partial compensation, with on average, 16.3 % vs. 10.6 % of the upward vs. downward shifts, but with high variability for the same talker between utterances and between talkers. However, this study was not designed to measure the delay in compensatory response. This delay was found in later studies to be around 160ms, at least when F1 is shifted upward (e.g. Tourville, Reilly, & Guenther, 2008), and when more complex spatial or temporal perturbations of formants trajectories are applied during the production of short sentences (Cai, Ghosh, Guenther, & Perkell, 2011). The smaller compensation of perturbation observed in studies involving unexpected perturbation compared to studies involving systematic perturbation, as well as the delay required to observe a compensatory response, confirm the idea that responses produced in the presence of the perturbation in adaptation studies are at least partially adaptive.

One of the most intriguing outcomes of Houde and Jordan (2002) was that the modification in formants was still present when talkers came back a month later to run a control study evaluating changes in production without perturbation. This long-term effect was attributed by the authors to implicit memory of the task or specific control mechanisms for

whispered speech. Although not reproduced in later work – as there was no study with equivalent long-term assessment in our review at least – Purcell and Munhall, (2006a) showed that 115 repetitions without perturbation after the training phase were not enough to fully return to the baseline state. The explanation introduced by Houde and Jordan echoes the idea that auditory-motor learning could be specific to some situations or ways of speaking, as discussed in generalization studies. The ability to memorize specific ways of speaking according to the situation could be a way to support fast speech adaptability in known situations. This idea could be further investigated by means of transfer of adaptation from one context to another as discussed below.

Finally, large inter-speaker variability was also pointed out in Houde and Jordan (2002) and then observed in all the subsequent studies. MacDonald et al. (2010, 2011) suggested that this variability is not clearly related to the variability in baseline production, nor to the size of the vowel space. Inter-speaker variability, as well as partial compensation, in formant adaptation studies was often discussed in terms of a tradeoff between auditory and somatosensory feedback. For example, Purcell and Munhall (2006a) suggested that “Some [speakers] may rely more on kinesthetic feedback and thus are not influenced as much by acoustic feedback” (p. 975), while Houde and Jordan (2002) suggested “it may be that there are differences across participants as to the degree to which they rely on auditory feedback” (p. 308). The tradeoff between auditory and somatosensory feedback, as well as the role of sensory acuity in adaptation was then explored in several papers, as described in the next section.

5.2. Perception acuity and sensory integration

Formant perturbations’ paradigms involve modifying the auditory feedback, i.e. sensory input of speech control system, and measuring the outcome in terms of speech production, or motor control. Hence these paradigms are by nature relevant to the question of the relationship between perception and production. Several aspects of this relationship have been investigated over the past two decades.

First, adaptation to auditory perturbations may be influenced by speakers’ sensory acuity. Auditory acuity has been positively correlated with adaptation magnitude in two studies (Martin et al., 2018;

Villacorta et al., 2007) involving 13 and 31 subjects respectively. Auditory acuity measurements were based on discrimination tasks in both cases. Villacorta et al. focused on acuity for F1 while Martin et al. measured acuity based on pitch and loudness, as well as melody discrimination tasks. A possible interpretation of the relation between adaptation magnitude and auditory acuity is that better acuity could lead speakers to have smaller goal regions for their production, resulting in higher adaptation (Villacorta et al., 2007). However, auditory feedback may not be the only feedback used to control speech production. Feng et al. (2011) investigated the relationship between the adaptation magnitude of F1 and the auditory acuity for F1, as well as somatosensory acuity for jaw position. They did not find a reliable correlation. However, fewer subjects were involved in this study than in previously cited ones (8 subjects vs. 13 and 31).

Feng et al. also combined a somatosensory perturbation induced by a robotic device pulling the jaw, with an auditory shift on F1. Using this procedure, they found that speakers mainly compensated for the auditory perturbation. They suggested that auditory feedback may be dominant over somatosensory input, but that their relative weight could evolve with speech experience. Using similar methods, Lametti, Nasir, and Ostry (2012) found that all speakers adapted for at least one of the two perturbations. The group who adapted to the somatosensory perturbation (half of the participants) did not significantly compensate for the auditory perturbation while the group that did not adapt to the jaw perturbation significantly compensated for the F1 shift. This observation suggests a speaker-specific sensory preference for either auditory or somatosensory inputs. In addition, the weights attributed to auditory and somatosensory feedback may vary according to the articulator (i.e. vocal folds, tongue or jaw) to control. Indeed, no correlation has been found in the magnitude of adaptation in F0, F1 and F2 across speakers while altering them simultaneously or separately (Eckey & MacDonald, 2015; MacDonald & Munhall, 2012). Interestingly, Trudeau-Fisette et al. (2017) showed that speakers with blindness adapted more to an F2 shift than control speakers, independently of their auditory acuity, and that they also produced larger articulatory changes in response to the auditory shift. Speakers with blindness may rely more on auditory feedback than control speakers, who

may have more precise somatosensory goals, probably built and supported by visual perception of speech.

However, sensory preference in the control of speech, which can be modeled by different weights attributed to each kind of sensory feedback, may also evolve with experience. Most studies on auditory-motor adaptation report a partial compensation for the auditory perturbation as already mentioned in the previous section. Some studies showed that the percentage of compensation relative to the magnitude of the perturbation decreases when the magnitude of the perturbation increases, reaching an asymptote, and can even tend to decrease for larger perturbations (Katseff & Houde, 2008; Katseff et al., 2012; MacDonald et al., 2010). Katseff et al. (2012) interpreted this phenomenon as evidence that the weights attributed to auditory and somatosensory feedback may vary according to experience: “For small discrepancies between auditory and somatosensory feedback, auditory feedback takes precedence, and for large discrepancies between auditory and somatosensory feedback, somatosensory feedback takes precedence” p. 307. Thus, a high-amplitude shift may lead the speech system to consider auditory feedback as unreliable and therefore give more weight to somatosensory feedback. In addition, the relative importance of sensory input may depend on the specific sounds produced. Several studies observed less compensation in closed vowels than in open vowels (Mitsuya et al., 2015; Purcell & Munhall, 2008; Reilly & Dougherty, 2013). This could be explained by better-specified somatosensory information in the former than in the latter case (Mitsuya et al., 2015). Another possible explanation is that the importance of F1 as an acoustic cue in perception may depend upon the vowel (Reilly & Dougherty, 2013).

5.3. Perceptual and phonological categories

Speech perceptual space is structured by phonological categories, which are delimited by perceptual boundaries. Niziolek and Guenther (2013) showed an effect of perceptual boundaries on the magnitude of compensation to unpredictable auditory perturbations. They observed that if the auditory signal resulting from the perturbation is near a boundary, the compensation, as well as the cortical activation, is higher than when it

is far from a boundary, the magnitude of the shift being equal. In addition, various studies have investigated the relation between perceptual boundary and adaptation to sustained auditory perturbations.

The influence of perceptual boundaries on adaptation can be investigated using perceptual learning on the perceptual contrast that is at stake in the adaptation paradigm. For instance, Shiller, Lametti, & Ostry (2013) manipulated speakers' perceptual boundaries between "head" and "had" through perceptual training preceding auditory-motor adaptation to a perturbation consisting of altering "head" into "had" (see Procedure P1p on Figure 3). The group whose boundary was shifted towards "head" (i.e. who was more likely to classify ambiguous stimuli as "had") adapted more to the auditory perturbation than the group whose boundary was shifted towards "had" by the perceptual training. Similarly, children adapted more to a perturbation transforming /beb/ into /bab/ after a perceptual training manipulating /ε/-/æ/ boundary towards /ε/ than before training (Shiller & Rochon, 2014). They also adapted more than children having undergone a perceptual training on an unrelated contrast. Furthermore, Lametti, Krol, et al. (2014) observed in adults that the amount of adaptation to auditory-feedback perturbation was correlated with the position of the perceptual boundary obtained through perceptual training.

Instead of using perceptual training, changes in perceptual boundaries were obtained by manipulating the pitch and formant of the carrier phrase "please say what this word is..." (Bourguignon, Baum, & Shiller, 2015, 2016). In this study, the group exposed to high carrier-phrase (high pitch and formants) had the boundary between 'bit' and 'bet' shifted toward 'bet'. They adapted more to an auditory feedback alteration transforming /ε/ into /i/ than the speakers exposed to low carrier-phrase (low pitch and formants). This finding suggests that "context-dependent plasticity in speech perception may also transfer to production" (Bourguignon et al., 2016, p. 1040). Interestingly, Bourguignon, Baum, and Shiller (2014) also showed an effect of the lexical status that can be interpreted in terms of perceptual boundaries. In their study, a group of speakers produced pseudo-words that resulted in real word when auditory perturbation was applied (e.g. "kess" changed into "kiss"). Another group produced real words that were transformed into pseudo-words by the same formant shift

(e.g. “less” changed into “liss”). The first group showed greater adaptation than the second group, indicating a lexical effect on auditory-motor adaptation.

The influence of phoneme categories on speech motor adaptation was also highlighted in cross-language studies. Mitsuya, MacDonald, Purcell, and Munhall (2011) contrasted the adaptation to upward and downward shifts in F1 in three groups: English speakers pronouncing “head”, Japanese speakers producing the Japanese word /he/ and Japanese speakers learning English, producing “head”. The magnitude of adaptation was equivalent in all groups in response to the downward shift, but the adaptation was smaller in Japanese than in English speakers in response to the upward shift. This difference is evidence for the influence of the phonological system in adaptation. Mitsuya, Samson, Ménard, and Munhall (2013) also showed differences between English speakers and French speakers in the adaptive response to the same auditory perturbation. In this study, a perception test suggested that this language effect on adaptation was mediated by a difference in perceptual boundaries: larger adaptation in French speakers was related to greater sensitivity to some phonetic contrasts.

Reciprocally, the influence of adaptation on perceptual boundaries has also been investigated. Lametti et al. (2014) incorporated perceptual tests in a classic auditory-motor procedure (Figure 3, procedure P1p), before and after the training phase – during which adaptation occurs – as well as after the after-effect phase, used here as a wash-out of adaptation. They observed that auditory-motor adaptation resulted in a shift of a perceptual boundary in the phonetic range of what speakers produced but not what speakers heard. For instance, speakers who produced “head” and heard an auditory feedback shifted toward “had”, compensated by producing an utterance closer to “hid”. Their perceptual boundary between “head” and “hid” was shifted toward “head”, that is, speakers became more likely to report hearing ‘hid’ in the perceptual test, while there was no effect on the perceived boundary between “head” and “had”. This result suggested that the change in perception was specifically driven by speech motor adaptation and not by the auditory input during learning. The interpretation of these results, together with the results of other studies on the effect of auditory-motor adaptation on categorical perception, was recently specified in a Bayesian modeling framework, suggesting

that speech motor adaptation results both in speech sound remapping and changes in phoneme categories (Patri, Perrier, Schwartz, & Diard, 2018). Yet, using a similar paradigm to that of Lametti et al. (2014), Schuerman, Meyer, and McQueen (2017) did not find significant influence of auditory-motor adaptation on related perceptual boundaries. It should be noted that this experiment had fewer subjects than Lametti et al. (2014); was run with speakers of Dutch as opposed to English; and used a continuum with isolated vowels rather than a continuum between words, during the perceptual test. However, this last study also recorded EEG signals during initial vs. final perception tests. The analysis of ERPs to the stimuli of the /ɛ/-/i/ continuum revealed changes in N1 and P2 components for ambiguous stimuli, which correlated with the magnitude of adaptation as measured by F1. The effect on both N1 and P2 suggest that auditory-motor adaptation influences both early perception and late perceptual decisions. Interestingly, Schuerman, Nagarajan, and Houde (2015) and Schuerman et al. (2017) showed that the adaptation to an auditory perturbation of F2 shifting the front vowel /i/ towards the back-vowel /u/ resulted in a shift in the perceptual boundary between “see” and “she”. More specifically, the shift in perceptual boundaries depended on the behavior of speakers during the adaptation task: speakers who followed the auditory perturbation had their perceptual boundary shifted in the opposite direction to that of speakers who compensated for the auditory feedback. This last group was more likely to categorize ambiguous stimuli as “see” than “she”, the place of articulation of the consonant /s/ being more anterior than /ʃ/. These findings are in agreement with the idea that some transfer of adaptation may occur between vowels and consonants articulated with a similar tongue position.

While this impact of a change in production on the perception of another contrast is actually a transfer from production to perception, the term *transfer* is typically investigated in speech production itself, from one utterance to another.

5.4. Transfer/specificity and speech units

In the limb movement literature, generalization of motor learning is the “ability to correctly extrapolate to contexts that are different from our limited experience” (Krakauer, Mazzoni, Ghazizadeh, Ravindran, &

Shadmehr, 2006, p. 1798). This extrapolation could be the result of an interpolation of previous experiences (Mattar & Ostry, 2007). Generalization has been extensively investigated in motor learning research, and in speech, in particular to address the specificity of motor adaptation and the underlying representations (Tremblay, Houle, & Ostry, 2008). Transfer of adaptation is usually defined as a positive generalization, as opposed to interference (Krakauer et al., 2006). However, we will use generalization or transfer to designate changes observed in untrained utterances after adaptive training, going in the same direction as adaptation. When no significant transfer is observed, changes related to adaptation are considered to be specific to the training utterance.

The investigation of generalization or transfer of adaptation relied on two different motivations. The first set of work focused on generalization as a way to assess the global vs. specific nature of auditory-motor mapping. This approach is derived from limb movement studies that analyzed generalization of visuomotor adaptation to address the global vs. specific nature of visuomotor mapping. The second set of work, that is sometimes an extension of the first one, considered generalization of auditory-motor learning as a way to assess the nature of speech production units, by questioning the linguistic level of auditory-motor mapping. This second approach was introduced by Houde and Jordan (1998) and is consistent with earlier work on transfer of perceptual learning to assess speech perception units (e.g. Chambers et al., 2010).

Different procedures were used to investigate generalization of auditory-motor adaptation. The first one is structured in “epochs” (Figure 3, P2). Each epoch includes utterances with feedback on and utterances with a masking noise, which can be either the training utterances or different utterances. Transfer is evaluated at the end of the training phase, when the perturbation is maximal by measuring changes in transfer (or test) utterances as compared with their baseline. Using this procedure, Houde and Jordan (1998) found significant transfer from the training words sharing the same vowel / ϵ / (“pep”, “peb”, “bep”, and “beb”), shifted toward / i / or / \ae / to the various test words (same vowel as training words – “gep”, “peg”, “teg”, or different vowels – “pip”, and “pap”). The amount of transfer was variable depending on the test word, but not statistically different. Consistent results were reported in Villacorta et al. (2007),

where adaptation on the vowel / ϵ / for nine CVC words to an F1 perturbation, significantly generalized to the same vowel in different CVC or to the vowels in “pit”, “pat”, and “pot”. Results were less consistent for “put” and “pete” and seemed to depend on the direction of the perturbation. Still with a similar procedure, but with perturbation of F1 trajectory in speakers of Mandarin, Cai et al., (2008) and Cai et al. (2010) found gradients of generalization that depended on the similarity in formant trajectory between a training triphthong and the tested utterances. Finally, Reilly and Pettibone (2017) tested generalization from the vowels /i/ vs. / ϵ / (embedded in a set of CVC utterances) to /i/, / ϵ / and / ϵ / (also in CVC) produced with a masking noise. In both training conditions, / ϵ / was the “near” vowel in test utterances, while /i/ and / ϵ / were either the same as the training vowel or the “far” vowels, depending on the training condition. Adapted speakers exhibited significant generalization to all vowels, regardless the training vowel. However, correlation between adaptation and generalization were unclear suggesting that generalization may depend on multiple factors and may be sensitive to inter-speaker variability.

Similar procedures, mixing training and transfer trials, were used in limb movement studies. However, the approach was later criticized. In particular, with this procedure, “the patterns of generalization observed are difficult to interpret, as transfer could reflect an averaging that takes place when subjects experience several training conditions simultaneously” (Rochet-Capellan et al., 2012 p. 1711). For this reason, other studies tested transfer after the training phase, when the feedback is turned-back to normal (Procedure P1t on Figure 3). In preliminary work, MacDonald et al. (2008) compared transfer tested in the course of training vs. after training. In both cases, speakers were trained on “head” shifted towards “had” and transfer was tested on the production of “hid” with unaltered feedback. When the transfer utterance “hid” was inserted during training, changes in “hid” were observed at the beginning of the training phase, but then its production came back to baseline. When tested after training, no change was observed at all in “hid”. Overall this suggests that adaptation is specific to the trained vowel, although it slightly depends on the training conditions. Pile, Dajani, Purcell, and Munhall (2007) then observed similar adaptation and lack of generalization toward “hid” or “hayed” (i.e. /hed/). Both studies were published in proceedings and were preliminary,

with restricted analyses. Later work by Rochet-Capellan et al. (2012) evaluated how adaptation to a perturbation of F1 in /pen/, /ben/, /ken/, /gen/, /ten/, /den/, /pan/, /pin/ then affect the production of /pen/ produced without perturbation. Results were consistent with previous work that tested generalization with a mixed procedure (Figure 3, P2): generalization was variable according to the training word and seemed to depend on the acoustical proximity between the training and the testing utterance. Another important result of this work was that the after-effect, assessed on the training utterance after the transfer phase, was still significant, suggesting that the production of the transfer utterance with normal feedback did not wash out adaptation. This last result is consistent with the idea that learning is related to the training experience, and at least to some extent specific to this experience.

Another way to assess specificity of adaptation is to evaluate how speakers can specifically compensate for several perturbations in the same training session. This approach is inspired by limb movement studies and in particular Osu, Hirai, Yoshioka, and Kawato (2004). In Rochet-Capellan and Ostry (2011), speakers produced “head” and “had” in random order with F1 shifted downward in “head” and upward in “had” and conversely (Procedure P3 in Figure 3). On average, speakers were able to change F1 frequency in opposite directions for “head” and “had”, suggesting that auditory-motor mapping is specific to each vowel. To assess whether auditory-motor mapping could be specific to a word, the authors then evaluated multiple adaptations for “head” and “bed” shifted in opposite directions and “ted” un-shifted. Again, on average, specific adaptation in opposite directions were observed for “head” and “bed” while F1 in “ted” remained unchanged, suggesting that different auditory-motor mappings could be built for a same vowel in different words. Similar results were obtained recently by Klein, Brunner and Hoole (in this book) with a Russian vowel in /d/ vs. /g/ CV syllables and a perturbation of F2. The authors also provided analysis of speakers’ data showing symmetrical vs. asymmetrical profiles of adaptation.

Altogether, these results suggest that generalization of auditory motor adaptation occurs in a way that depends on the similarity between the training and the testing utterance and that specific control can be achieved, at least under specific conditions. The results were interpreted as an

indication of global control for vowel production vs. specific control. Furthermore, generalization from a vowel to the same vowel in different contexts suggests that auditory-motor mapping could occur at the level of the phoneme. It is thus a way to question the structure of feedforward mapping, and the nature of its underlying representations (Houde & Jordan, 1998). The fact that transfer is in general smaller than after-effect suggests that word context may play a role. The idea that multiple representations may coexist in auditory-motor mapping of speech was directly assessed in recent papers by Caudrelier et al. (Caudrelier et al., 2016; Caudrelier et al., 2018). In this work, several linguistic levels were contrasted by assessing transfer on test utterances that shared either the same vowel, and/or the same syllable or was the same word as the training utterance. Transfer was smaller (although significant) at the vowel level than transfer to the same syllable, which was lower than after-effect in the same word, suggesting that these three levels – words, syllables, phonemes – could coexist in parallel in the structure of the speech sound map. This conclusion is consistent with multiple traces connectionist models of long-term memory (Ans, Carbonnel, & Valdois, 1998; Carbonnel, Charnallet, & Moreaud, 2010) in the sense that multiple units could emerge as common information of multiple experiences (Goldinger, 1998; Hintzman, 1986). Specific production of the vowel to the syllable or word context also questions the role of coarticulation in adaptation and transfer of adaptation, a topic introduced in a preliminary paper by Berry et al. (2014).

In addition to the theoretical insights mentioned above, a better understanding of generalization in speech may have clinical implications in speech rehabilitation (e.g. after stroke), since transfer from training with a speech therapist to daily life is essential (Aichert & Ziegler, 2013). Other clinical applications are described in the next section.

5.5. Pathology affecting speech production

Auditory feedback perturbation paradigms may be instrumental in the understanding of mechanisms underlying disorders related to or affecting speech production. In particular, low compensation or adaptation observed in patients with a given pathology is regarded as evidence for a lack of sensorimotor integration or as an impairment of feedforward control mechanisms.

Stuttering is suspected to be driven by abnormal integration of sensory input in speech motor control, and has been an early target for auditory perturbation studies, and more recently for studies using formant perturbations. Cai et al. (2012) observed smaller compensation to unpredictable perturbation of formants in persons who stutter compared to control participants. The latency of compensation was however found to be equivalent in both groups. According to the authors, this suggests impairment of the inverse model responsible for translating auditory error detection into proper correction in motor commands. Reduced responses to formant perturbations were also observed in adaptation studies, with systematic perturbations. Sengupta, Shah, Gore, Loucks, & Nasir (2016) found smaller adaptation in adults who stutter as compared with control speakers that was also related to anomalous EEG phase coherence. This hints at a miscommunication between speech sensory and motor areas, which confirms a potential deficit in sensorimotor integration in people who stutter. A recent study, Daliri, Wieland, Cai, Guenther, & Chang (2018) also found reduced adaptation in adults who stutter compared to control speakers. However, the difference was not observed in children who stutter as compared with their aged-match controls. These results suggest that reduced adaptation observed in adults may be a consequence of compensatory strategies induced by the pathology rather than a root cause.

Terband, Van Brenk, and van Doornik-van der Zee (2014) used a similar adaptation paradigm as Daliri et al. (2018) with children with CAS (Childhood Apraxia of Speech). CAS was described as “a disordered development of the functional synergies/coordinative structures that underlie speech motor coordination causing impairment of the forward model leading to poor feedforward control” (Terband et al., 2014, p. 66). In agreement with this description, children with CAS were shown to follow the auditory perturbation on average, while their aged-match controls adapted to the perturbation by compensating for it.

Van den Bunt et al. (2017) used formant adaptation to assess the nature of the phonological deficit observed in dyslexia, known as a “difficulty in acquiring fluent word-decoding skills” (p. 1). Adults with dyslexia showed greater adaptation and after-effects than control speakers to a formant feedback perturbation that doesn't cross a phonemic boundary (i.e. an

allophonic perturbation). Moreover, a negative correlation was observed between reading skills and the magnitude of adaptation: the worse the reading score, the larger the adaptation. This result could be interpreted as a weaker perceptual magnet effect (Kuhl et al., 2008) in speakers with dyslexia and supports theories claiming that dyslexia is associated with a greater distinction between allophones, which may lead phoneme categories to be less prominent. However, a condition with a perturbation crossing the phonetic boundary is required to further support this hypothesis.

Compensation or adaptation to formant perturbations were also investigated in populations with neurogenetic or neurodegenerative diseases. Demopoulos et al. (2018) used adaptation to formant perturbation to address the origin of the speech production deficit observed in young individuals with a subtype of autism (due to a 16p11.2 deletion). The adaptation was reduced in this population as compared with age-matched controls while compensation to unexpected perturbation of F0 was larger. According to the authors, this suggests that feedforward models could be altered in people with 16p11.2 deletion, leading to an over-reliance on feedback control. A comparable profile of larger compensation to unexpected perturbation of F0 was observed in patients with Parkinson Disease (PD). However, both compensation and adaptation to unexpected vs. constant formant perturbation were reduced in speakers with PD as compared with age-matched control speakers (Mollaei, Shiller, Baum, & Gracco, 2016; Mollaei, Shiller, & Gracco, 2013). The authors interpreted the difference in pitch and formant compensation in terms of somatosensory and muscle activation deficits of the larynx and oral cavity. This dissociation between compensation to F0 vs. formant perturbations calls into question the conclusion of Demopoulos et al. (2018): as feedback control was only assessed with F0 in speakers with 16p11.2 deletion, it remains unclear whether they indeed rely more on feedback control in general or if the effect was specific to F0 control. Finally, Parrell, Agnew, Nagarajan, Houde, & Ivry (2017) found that speakers with cerebellum degeneration compensate for unexpected formant perturbations more than their age-matched controls, while they show weaker adaptation to sustained perturbation. This suggests that the cerebellum plays an important role in feedforward control, and probably less in feedback control. The involvement of the cerebellum in feedback control is discussed in the next section.

5.6. Neural basis of speech motor learning

The neural correlates of speech motor control and learning have been investigated through a variety of techniques, including EEG, fMRI, rTMS and tDCS.

fMRI is not suitable to observe changes in the timeframe of adaptation to sustained perturbation because it could be confounded with low-frequency noise observed in fMRI (Zheng et al., (2013)). However, it is feasible to investigate the neural networks involved in feedback control using unpredictable perturbations. In Tourville et al. (2008), trials under altered auditory feedback (as opposed to normal feedback) were associated with increased bilateral activation in posterior auditory cortex (including posterior Superior Temporal Gyrus, pSTG, and Planum Temporale, PT). This observation is regarded as evidence for the existence of auditory error cells, dedicated to detect errors in auditory feedback. The increased activation in right pSTG was observed to be enhanced when auditory perturbation outcomes were close to a perceptual boundary. In addition, Tourville et al., (2008) found increased right activation in ventral Motor and Premotor Cortex (vMC and vPMC, respectively) and anterior medial cerebellum (amCB). This suggests that feedback control involves mainly the right hemisphere whereas the left hemisphere, which is known to be dominant in speech production, would be mainly associated with feedforward control. Zheng et al., (2013) conducted further fMRI investigation. Their experimental procedure consisted of production trials with normal feedback, altered feedback (with F1 shift) and feedback with masking noise. Speakers then passively listened to every signal corresponding to their auditory feedback in the production session. Combining fMRI with an analysis of neural pattern similarity analysis enabled differentiation of three functional networks: an error signal network (including right AG, right SMA, and bilateral cerebellum), a passive listening network, and a network responding to both production and passive listening conditions, that may correspond to sensorimotor integration, located in bilateral Inferior Frontal Gyrus (IFG).

The Inferior Parietal Lobe (IPL), which comprises Supramarginal Gyrus (SMG) and Angular Gyrus (AG) may be involved in multisensory integration. An rTMS stimulation applied over the SMG just before the auditory-motor adaptation procedure reduced adaptation responses in

comparison with a sham stimulated group (Shum et al., 2011). Similarly, a tDCS stimulation applied over IPL affected auditory-motor adaptation (Deroche, Nguyen, & Gracco, 2017). More specifically, anodal stimulation aiming at facilitating neuronal excitability resulted in stronger adaptation magnitude whereas cathodal stimulation, which has an inhibitory effect, prevents auditory-motor adaptation to predictable perturbations.

Lametti, Smith, Freidin, and Watkins (2018) investigated the specific role of two areas involved in motor control, the cerebellum and the premotor cortex. In this experiment, anodal tDCS was applied during the baseline phase and the training. The auditory perturbation consists of an F1 shift making the training words “bed”, “head” and “dead” sound more like “bad”, “had” or “dad”, respectively. Stimulations over either motor cortex or cerebellum were both found to lead to higher adaptation and/or after-effect than in the sham-stimulated group. Interestingly, stimulation over the cerebellum increased error compensation on F1, while stimulation of the motor cortex also led to adaptation in F2. Adaptation in F2 when altering F1 only has been reported for the front vowel / ϵ / with variable size-effects (MacDonald et al., 2011; Rochet-Capellan & Ostry, 2011; Villacorta et al., 2007). Changing F2 in answer to a perturbation of F1 may be a strategy to reach an appropriate phoneme auditory category, as F1 and F2 vary at the same time in the contrast of front vowels. Thus, the cerebellum is suggested to contribute to error correction only, while motor cortex may lead to more general adaptation, possibly related to previously learnt movements.

While rTMS and tDCS can reveal the functional role of a specific brain area, neuronal oscillations as observed in EEG combined with phase coherence analysis may provide insights into the communication between brain areas as proposed by Sengupta and Nasir (2015). Phase coherence over a specific brain area can also represent a measure of this area’s engagement. In this study, a redistribution of phase coherence in specific frequency bands (theta and gamma bands) occurred at the end of the training phase and was related to the amount of speakers’ adaptation. This phenomenon was interpreted as a sign of the establishment of a new feedforward map (i.e. associating an auditory target to a motor gesture that enables the speaker to reach it) together with increased engagement of sensorimotor areas. Sengupta and Nasir (2016) then found that by late training, power in specific frequency bands during speech planning and

speech production was related to whether speakers were adapting to the auditory perturbation or not. Finally, Sato and Shiller (2018) analyzed event-related potentials (ERPs) during adaptation to an increase of F1. They observed that electro-cortical potentials at certain temporal windows (N1, P2) amplitude mirrors adaptation, as larger adaptation magnitude correlated with smaller N1/P2 amplitude. This larger speaking-induced suppression with learning was interpreted as an indication of auditory prediction during speaking.

5.7. Speech development

Auditory perturbation is an artificial way to generate speech learning, which otherwise occurs in *natural* situations: learning a new language, as well as during the development of speech. Studying adaptation to perturbations in typical adult speakers might help understand potential mechanisms occurring in these natural situations. It also questions the way children learn speech sounds. Daliri et al. (2018) and Terband et al. (2014) studied adaptation in atypical development, as reported in the “Pathology” section. Shiller and Rochon (2014) investigated the relation between adaptation on perceptual boundaries in children, as reported in the “*Perceptual and phonological categories*” section. MacDonald, Johnson, Forsythe, Plante, and Munhall (2012) and Terband and Van Brenk (2015) focused on adaptation in typically developing children at different ages. Terband and Van Brenk (2015) found greater adaptation in 4 to 9-year-old children than in adults, although the magnitude of adaptation did not correlate with age in the group of children, and the proportion of children exhibiting a consistent compensatory response was lower than in adults. MacDonald et al. (2012) showed that 4-year-old children adapted to a sustained perturbation with a similar magnitude of adaptation as adults, whereas 2-year-old toddlers did not adapt at all. This could suggest that toddlers ignore their own auditory feedback to focus on external stimulation or have an immature feedforward control. According to Messum and Howard (2012), this observation contradicts the widely held view that children learn speech sounds by imitation, which would require them to listen to what they produce and try to make it match what they want to imitate. Instead, it supports the idea that a child learns to speak thanks to a tutor: “Mothers reflect (or mirror) what

their children say, but such imitation generally takes the form of reformulation into well-formed sounds of the ambient language, rather than simple mimicry” (Messum & Howard, 2012, p. 160). Thus, plasticity observed in adults in the situation of adaptation to auditory perturbations may be different in nature to what occurs in the early speech development.

5.8. Surface effects and speakers’ characteristics

Other effects related to speakers or context, like the characteristics of the prompt during the adaptation procedure, may influence speech adaptation. Alsius, Mitsuya, Latif, and Munhall (2017) investigated the influence of the stimulus used to prompt the training word “head” by contrasting visual and auditory modalities as well as linguistic vs non-linguistic prompts. No effect of the sensory modality was found on the magnitude of adaptation but linguistic prompts (“head” as a spoken or written word) were found to induce more adaptation than non-linguistic prompts (a cross or a tune). Similarly Sato and Shiller, (2018) found no difference in the magnitude of adaptation between visual and auditory modalities. In addition, Caudrelier et al. (2018) investigated whether naming a picture or reading a word aloud would make a difference in adaptation and in transfer. Although no effect was found in the adaptation response, the pattern of generalization was influenced by the prompt used during the transfer phase, regardless of the training prompt, hinting at possible surface effects.

With regards to speakers’ abilities, Martin et al. (2018) found no correlation between general executive control and adaptation magnitude. In a preliminary study, Dimov, Katseff, and Johnson, (2012) investigated the influence of speakers’ characteristics including some social and personal aspects. In particular, less empowered subjects were found to adapt more than more empowered ones. Finally, Munhall et al. (2009) reported equivalent adaptation in naïve speakers and in speakers who were informed of the shift and who were asked to compensate or not. These results suggest that auditory-motor recalibration is at least in part an automatic process. More work is required to better understand the complexity of adaptive profiles that might be determined by numerous factors, as discussed in the next section.

6. Research outlook on formant perturbations

In this section, we identify some perspectives for future studies in adaptation to formant perturbations, in relation to methodological aspects as well as to some of the reviewed research questions.

6.1. Toward standards to investigate and report adaptation to formant perturbations

Various interests have motivated adaptation to formant perturbations studies in various teams. This induced the use of different methods to alter formants but also different procedures and analyses. These methodological differences often make studies difficult to compare directly. Therefore some standards should be developed, in particular to facilitate meta-analyses of formants perturbations studies, at least with regards to the way to report the methods and the results. Munhall, Purcell and collaborators studies are very interesting in this regard, as they have involved a significant number of speakers and have used similar methods to alter formants to run the adaptation and to analyze the data. A number of questions should be taken into consideration when designing and reporting studies. Some of them may also require further methodological studies, in line with Munhall and collaborators work. For instance:

- Should the participants be only females or males? What is the effect of mixing vs. not mixing gender on adaptation?
- This first question could be crossed with the effect of the type of perturbation: should the perturbation be absolute vs. relative, formant values being clearly different across gender? What is the effect of shifting only F1 vs. F1 and F2 in opposite direction?
- Whether participants are monolingual or multilingual should be controlled and reported, and as far as possible kept available for meta-analysis. Indeed, adaptation seems influenced by perceptual categories, which are related to phonological systems of languages. One of the best ways to address the question would be to be able to compare large datasets recorded around the world in the different research topics.
- What is the real effect of bone conduction on adaptation? This question has not been addressed systematically, although it has been considered in the conception of apparatuses to shift formants. Most studies used

quite high sound intensity of feedback and/or mixed the signal with noise. The effect of the feedback level, the signal to noise ratio as well as the type of noise on adaptation were not systematically reported.

- What is the real effect of the perturbation on the signal heard by speakers? This question is rarely investigated in papers, while the obtained perturbation can be far from the expected one (Mitsuya et al., 2015). In particular, when using existing packages such as Audapter, delay in feedback should be checked, as it could depend on the properties of the OS and computer hardware. The evaluation of formants provided by the tool, especially when applying unusual shifts, should be verified as there is no guarantee that the system will be able to track and shift the formants in the expected way. This is true for all the systems and could be easily verified by comparing the obtained formant values with corresponding spectrograms or with values assessed by an independent formant assessment software. This approach was used in Reilly and Pettibone (2017).
- Due to the high variability in adaptation magnitude between participants, apparent differences on some parameters of adaptation between conditions are often found to be non-significant. Some effects and, in particular, surface effects such as visual vs. audio prompts might exist but may require testing a large number of speakers to reach significance. This could also be the case for effects related to the direction of the perturbation or to the number of trials during the hold phase as well as to the way the perturbation is introduced. At the very least, non-significant results between different groups of speakers should be interpreted carefully, in relation to this large variability.

These examples suggest that methodological aspects should be directly addressed and clearly reported to help teams working in the field share standards and enable the constitution of large databases. Large between-subjects variability suggests that adaptation to formant perturbations is a complex phenomenon, influenced by different factors. Multifactorial analyses such as introduced in Dimov et al. (2012) could be run on large datasets, but this requires – at the very least – recording of systematic information about the participants and reporting clear information about the perturbation and its real effect.

6.2. Topics which will benefit from further investigation

Due to the broad range of research topics addressed by formant perturbation studies, more studies are still required to reproduce or better understand some results. This is particularly the case for the effect of adaptation on categorical perception, as results between studies have been sometimes inconsistent. Only a few studies were published on the effect of adaptation to formant perturbations on categorical perception of speech (Lametti et al., 2014; Schuerman et al., 2015; Schuerman et al., 2017; Schuerman, Nagarajan, et al., 2017) with some inconsistent findings between Lametti et al. (2014) and Schuerman et al. (2017). The two studies were run with speakers of different languages (English vs. Dutch) and with different types of continua for the perceptual test (words vs. vowels). It would be useful to gain more awareness of other attempts with non-significant or inconsistent profiles of perceptual changes following adaptation if any exist. This will avoid a publication bias towards significant-only results that seems to be a sensitive topic for this research question, in particular as the effects of speech production on changes of categorical boundaries may be sensitive to numerous variables, including the number of speakers, their gender, regional accent, languages skills etc. Replication is also required as the involvement of the motor system in perception is an important challenge for speech research more generally.

Investigating the development of feedback and feedforward control systems and their potential interaction in typically developing children is also an important topic to further develop using formant perturbations paradigm. Moreover, using compensation to unpredictable perturbations in conjunction with sustained perturbations in atypical speakers may shed light on the root causes of some pathologies affecting speech production. For instance, van den Bunt et al. (2017) provides a rather convincing explanation about the sensorimotor bases of dyslexia, which could be further investigated in children. As adaptation has been shown to interact with phoneme categories, it allows investigating the development of phonological categories in both typical children and children with phonological disorders.

An important topic also under-investigated so far is the influence of extraneous factors (i.e. not directly related to language or speech) on

auditory-motor adaptation. First results by Munhall et al. (2009) suggested that the magnitude of the compensation is relatively independent of the awareness of the experimental aim and that speakers compensate even when asked not to compensate. This suggests that adaptation is quite independent from higher cognitive functions such as attention. Martin et al. (2018) also found no significant contribution of general executive control skills on adaptation. However, the preliminary work by Dimov et al. (2012) suggests that variables related to speakers' social status may play a role. Further investigations linking working memory abilities, attention levels etc. to formant adaptation will help tackle the mainstream issue of the link between cognitive and sensorimotor functions. This topic, as a number of others, has already been investigated in adaptation or compensation to F0 perturbation (Guo et al., 2017; Hu et al., 2015; Scheerer, Tumber, & Jones, 2015). Last but not least, results by MacDonald et al. (2012) showing that toddlers do not adapt and the associated discussion of this result by Messum and Howard (2012) suggest that the communicative context may also influence adaptation. The question was investigated in birds by Sakata and Brainard (2009) suggesting larger adaptation when the song is produced in presence of another bird but also in humans with other type of perturbations such as speech in noise (Garnier et al., 2010). Social context might thus be relevant to question the real nature of speech targets.

An important topic not developed in this chapter is a systematic analysis of the results of formant perturbation studies in relation to current models of speech production. A joint analysis with the results of other auditory and somatosensory perturbation studies could improve our understanding of feedback and feedforward controls.

Finally, as it is relevant to the link between learning and memory, we would like to emphasize that transfer of adaptation was under-studied so far, despite its potential to bring insight into the nature of speech representations. As already introduced in Houde and Jordan (1998), transfer of learning is an empirical tool to question the nature of speech production units. This approach should be better connected to the equivalent approach developed for perceptual learning (e.g. Chambers et al., 2010). As noted by Cai et al. (2010) patterns of transfer question the way models of speech production represent sensorimotor mapping: both significant

generalization effect, as well as gradient effects should be explained. These models should also be adapted to integrate results from transfer or multiple adaptation studies suggesting that the mapping between auditory and articulatory domains could occur at different linguistic levels and be related in some way to the training word. But more generally, adaptation might be related to the episode of learning, as also discussed in Houde and Jordan (2002) when explaining the long-term effects of adaptation by implicit memory. We strongly believe that understanding the link between sensorimotor learning and memory would be a fruitful path towards understanding of embodied cognition and the links between language and speech. In any case, identifying the condition of specificity vs. generalization of adaptation will clearly contribute to the debate on the nature of speech production representation and to the debate on the nature of internal models and their relation to sensorimotor memories.

7. Conclusion

Twenty years ago, Houde and Jordan introduced formant perturbations in auditory feedback as a new paradigm to explore speech production. This seminal study is cited by papers in various domains: speech production and perception in general, studies using other kind of perturbations related to speech (e.g. pitch alteration, vocal tract perturbation), motor control as well as vocalizations in animals. Moreover, it has inspired a whole research field which is still in expansion. In this review, we scanned all studies citing Houde and Jordan (1998, 2002) and selected 77 articles focused on formant perturbations. The perturbation systems designed for this purpose are reported and described in the review. The main research topics addressed in these studies are also explained, along with their main findings.

The formant perturbation paradigm proved to be insightful in exploring the relationship between speech production and perception. First, the observation of responses to auditory perturbations has shed light on the role of auditory feedback in speech production, and the mechanisms that control it. Experimental findings have been incorporated in speech production models, although some results still need to be modeled. Altering both auditory and somatosensory feedback showed that both modalities

are integrated in the control of speech in a manner that may be specific to the speaker and/or to the task (e.g. the vowel to produce). Associating perceptual categorization tasks and training with formant perturbations revealed a close relationship and mutual influence between speech motor control and phonological categories, mediated by categorical perception.

This relationship between motor control and linguistic units (e.g. phonemes or syllables) has also been explored by observing the generalization of auditory-motor adaptation. Generalization, or transfer, has been observed from a vowel to the same vowel in different words, suggesting the existence of an underlying phoneme representation. While transfer may occur from one vowel to another, supporting the idea of broad generalization in speech learning, the magnitude of transfer seems to depend on some similarity relationship between the training and the transfer utterances. Moreover, simultaneous adaptation to opposite perturbations has been observed in two different vowels and even in the same vowel in different words. This apparent contradiction may represent a challenge for speech production models, as it requires much flexibility in the translation of auditory goals into articulatory gestures, and questions the nature of mental representations interfacing with speech articulation.

Studies in cognitive neurosciences have pinpointed neural correlates of sensory integration and motor control in speech production, in terms of brain regions as well as communication networks and frequency bands. While studying patients with cerebellar degeneration also contributes to this purpose, research in other pathologies, including stuttering, Parkinson's disease, dyslexia, developmental speech disorders, and some autism subtypes, have benefited from formant perturbation experiments in understanding of the main causes and mechanisms underlying these specific disorders. Finally, studying compensation and adaptation in children gives insights in the development of sensorimotor processes at stake in speech production. Effects of communicative situation or social context may also be explored, as it has proven influential in some speech motor control characteristics in adults. Further investigations in children in various communicative contexts could eventually shed light on one of the most intriguing questions in our research field: how does a child learn to speak?

Beyond these core topics associated with Houde and Jordan's paradigm, other questions have emerged in relation to speakers' cognitive functions and social characteristics, as well as learning context. The prompt has been suggested to influence adaptation and transfer pattern. Moreover, Houde and Jordan had already noticed that adaptation was still there when speakers were tested one month later with normal feedback. This observation may suggest that learning is to some extent specific to the context in which it occurs, the testing room for instance. This is consistent with multiple-trace memory models or exemplar-based views (Goldinger, 1998; Hintzman, 1986), according to which each event is recorded in the brain in the form of a trace combining multiples elements from sensory inputs. Being confronted with one of these elements may activate all the traces containing it, and therefore the other elements associated with it. Thus, the specific context of the testing room may reactivate the adaptation that had washed out in other contexts. Investigating retention of adaptation in various time ranges and contexts may pave the way to fruitful research exploring the relationship between speech, learning and memory.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007–2013 Grant Agreement no. 339152).

References

- Aichert, I., & Ziegler, W. (2013). Segments and syllables in the treatment of apraxia of speech: An investigation of learning and transfer effects. *Aphasiology*, 27(10), 1180–1199.
- Alsius, A., Mitsuya, T., Latif, N., & Munhall, K. G. (2017). Linguistic initiation signals increase auditory feedback error correction. *The Journal of the Acoustical Society of America*, 142(2), 838–845.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, 105(4), 678–723.
- Berry, J.J., Jaeger, I.V., Wiedenhoft, M., Bernal, B.A., & Johnson, M.T. (2014). Consonant context effects on vowel sensorimotor adaptation.

In: ISCA (Ed.): *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, (pp. 2006–2010).

- Berry, J.J., North, C., & Johnson, M.T. (2014). Sensorimotor adaptation of speech using real-time articulatory resynthesis. In *IEEE Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 3196–3200).
- Berry, J.J., North, C., Meyers, B., & Johnson, M.T. (2013). Speech sensorimotor learning through a virtual vocal tract. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060099). ASA.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, 8, 208.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2015). Extrinsic talker normalization alters self-perception during speech. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (Eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2016). Please say what this word is—Vowel-extrinsic normalization in the sensorimotor control of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 42(7), 1039–1047.
- Brainard, M.S., & Doupe, A.J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, 1(1), 31–40.
- Brumberg, J.S., Krusienski, D.J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A.L., & Schalk, G. (2016). Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *PloS one*, 11(11), e0166872.
- Burnett, T.A., & Larson, C.R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *The Journal of the Acoustical Society of America*, 112(3), 1058–1063.
- Cai, S., Beal, D.S., Ghosh, S.S., Tiede, M.K., Guenther, F.H., & Perkell, J.S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation. *PloS one*, 7(7), e41830.

- Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong/iau. *Proceedings of the 8th International Seminar on Speech Production (ISSP)*, (pp 65–68).
- Cai, S., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong/iau/and its pattern of generalization. *The Journal of the Acoustical Society of America*, 128(4), 2033–2048.
- Cai, S., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience*, 31(45), 16483–16490.
- Carbonnel, S., Charnallet, A., & Moreaud, O. (2010). Organisation des connaissances sémantiques: des modèles classiques aux modèles non abstraits. *Revue de Neuropsychologie*, 2(1), 22–30.
- Casserly, E.D. (2015). Effects of real-time cochlear implant simulation on speech production. *The Journal of the Acoustical Society of America*, 137(5), 2791–2800.
- Caudrelier, T., Perrier, P., Schwartz, J., Rochet-Capellan, A. (2018) Picture naming or word reading: Does the modality affect speech motor adaptation and its transfer? *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2018)*, (pp. 956–960).
- Caudrelier, T., Perrier, P., Schwartz, J.-L., & Rochet-Capellan, A. (2016). Does auditory-motor learning of speech transfer from the CV syllable to the CVCV word? *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)*, (pp. 2095–2099).
- Caudrelier, T., Schwartz, J.-L., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does it tell us about speech production units? *Journal of Speech, Language, and Hearing Research*, 61(7), 1613–1625.
- Chambers, K.E., Onishi, K.H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 821–828.

- Chon, H., Kraft, S.J., Zhang, J., Loucks, T., & Ambrose, N.G. (2013). Individual variability in delayed auditory feedback effects on speech fluency and rate in normally fluent adults. *Journal of Speech, Language, and Hearing Research*, 56(2), 489–504.
- Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human brain mapping*, 9(4), 183–191.
- Daliri, A., Wieland, E.A., Cai, S., Guenther, F.H., & Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental science*, 21(2), e12521.
- de Bruijn, M.J., ten Bosch, L., Kuik, D.J., Witte, B.I., Langendijk, J.A., Leemans, C.R., & Verdonck-de Leeuw, I.M. (2012). Acoustic-phonetic and artificial neural network feature analysis to assess speech quality of stop consonants produced by patients treated for oral or oropharyngeal cancer. *Speech Communication*, 54(5), 632–640.
- Demopoulos, C., Kothare, H., Mizuiri, D., Henderson-Sabes, J., Fregeau, B., Tjernagel, J., Houde, J.F., Sherr, E.H., & Nagarajan, S.S. (2018). Abnormal speech motor control in individuals with 16p11.2 deletions. *Scientific Reports*, 8(1), 1274.
- Deroche, M.L., Nguyen, D., & Gracco, V.L. (2017). Modulation of speech motor learning with transcranial direct current stimulation of the inferior parietal lobe. *Frontiers in Integrative Neuroscience*, 11, 35.
- Dimov, S., Katseff, S., & Johnson, K. (2012). Social and personality variables in compensation for altered auditory feedback. *UC Berkeley PhonLab Annual Report*, (6)6, pp. 259–282.
- Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22(1), 567–631.
- Eckey, A., & MacDonald, E. (2015). Compensations of F0 and formant frequencies in a real-time pitch-perturbation paradigm. *Fortschritte der Akustik DAGA'15*, (pp. 1444–1447).
- Eliades, S.J., & Miller, C.T. (2017). Marmoset vocal communication: Behavior and neurobiology. *Developmental Neurobiology*, 77(3), 286–299.

- Feng, Y., Gracco, V.L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679.
- Frank, A.F. (2011). *Integrating linguistic, motor, and perceptual information in language production*. Doctoral dissertation, University of Rochester.
- Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608.
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.
- Golfinopoulos, E., Tourville, J.A., & Guenther, F.H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*, *52*(3), 862–874.
- Grimme, B., Fuchs, S., Perrier, P., & Schöner, G. (2011). Limb versus speech motor control: A conceptual review. *Motor Control*, *15*(1), 5–33.
- Guo, Z., Wu, X., Li, W., Jones, J.A., Yan, N., Sheft, S., Liu, P., & Liu, H. (2017). Top-down modulation of auditory-motor integration during speech production: The role of working memory. *Journal of Neuroscience*, *37*(43), 10323–10333.
- Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213*(5), 84–97.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67–99.
- Hintzman, D. L. (1986). «Schema abstraction» in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428.
- Houde, J.F. (1997). *Sensorimotor adaptation in speech production*. Doctoral dissertation, MIT.
- Houde, J.F., & Chang, E.F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, *33*, 174–181.
- Houde, J.F., & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216.

- Houde, J.F., & Jordan, M.I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310.
- Hu, H., Liu, Y., Guo, Z., Li, W., Liu, P., Chen, S., & Liu, H. (2015). Attention modulates cortical processing of pitch feedback errors in voice control. *Scientific Reports*, 5, 7812.
- Hubl, D., Schneider, R. C., Kottlow, M., Kindler, J., Strik, W., Dierks, T., & Koenig, T. (2014). Agency and ownership are independent components of ‘sensing the self’ in the auditory-verbal domain. *Brain topography*, 27(5), 672–682.
- Jones, J.A., & Munhall, K.G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251.
- Jones, J.A., & Munhall, K.G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, 113(1), 532–543.
- Katseff, S., & Houde, J. (2008). Partial compensation in speech adaptation. *UC Berkeley Phonology Lab Annual Reports*, 4(4), (pp. 445–461).
- Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A trade-off with somatosensory feedback? *Language and Speech*, 55(2), 295–308.
- Kelso, J.S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C.A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812–832.
- Klein, E., Brunner, J., & Hoole, P. (2019). Spatial and temporal variability of corrective speech movements as revealed by vowel formants during sensorimotor learning. In S. Fuchs, J. Cleland & A. Rochet-Capellan (eds.) *Speech production and perception: Learning and memory*. Peter Lang Publisher (*current book*).
- Krakauer, J.W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., & Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS biology*, 4(10), e316.

- Krakauer, J.W., Pine, Z.M., Ghilardi, M.-F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, *20*(23), 8916–8924.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, *363*, 979–1000.
- Lametti, D.R., Krol, S.A., Shiller, D.M., & Ostry, D.J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning: *Psychological Science*, *25*(7), 1325–1336.
- Lametti, D.R., Nasir, S.M., & Ostry, D.J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*(27), 9351–9358.
- Lametti, D.R., Rochet-Capellan, A., Neufeld, E., Shiller, D.M., & Ostry, D.J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, *34*(31), 10339–10346.
- Lametti, D.R., Smith, H.J., Freidin, P.F., & Watkins, K.E. (2018). Cortico-cerebellar networks drive sensorimotor learning in speech. *Journal of Cognitive Neuroscience*, *30*(4), 540–551.
- Lane, H., Matthies, M.L., Guenther, F.H., Denny, M., Perkell, J.S., Stockmann, E., Tiede, M., & Zandipour, M. (2007). Effects of short-and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research*, *50*(4), 913–927.
- Li, W., Chen, Z., Yan, N., Jones, J. A., Guo, Z., Huang, X., Cheng, S., Liu, P., & Liu, H. (2016). Temporal lobe epilepsy alters auditory-motor integration for voice control. *Scientific reports*, *6*, 28909.
- Maas, E., Mailend, M.-L., & Guenther, F.H. (2015). Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language, and Hearing Research*, *58*(2), 185–200.
- Maas, E., Robin, D.A., Hula, S.N.A., Freedman, S.E., Wulf, G., Ballard, K.J., & Schmidt, R.A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298.

- MacDonald, E.N., Goldberg, R., & Munhall, K.G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, 127(2), 1059–1068.
- MacDonald, E.N., Johnson, E.K., Forsythe, J., Plante, P., & Munhall, K.G. (2012). Children's development of self-regulation in speech production. *Current Biology*, 22(2), 113–117.
- MacDonald, E.N., & Munhall, K.G. (2012). A preliminary study of individual responses to real-time pitch and formant perturbations. *The Listening Talker: An interdisciplinary workshop on natural and synthetic modification of speech in response to listening conditions*. 2012, (pp. 32–35).
- MacDonald, E.N., Pile, E., Dajani, H., & Munhall, K.G. (2008). The specificity of adaptation to real-time formant shifting. *Proceedings of the International Seminar on Speech Production, 2008*, (pp. 397–400).
- MacDonald, E.N., Purcell, D.W., & Munhall, K.G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, 129(2), 955–965.
- Martin, C.D., Niziolek, C.A., Duñabeitia, J.A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J.F. (2018). Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience*, 12, 91.
- Mattar, A.A., & Ostry, D.J. (2007). Modifiability of generalization in dynamics learning. *Journal of Neurophysiology*, 98(6), 3321–3329.
- Max, L., & Maffett, D.G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience letters*, 591, 25–29.
- Max, L., Wallace, M.E., & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. *Proceedings of the 15th International Congress of Phonetic Sciences*, Futurgraphic Barcelona, Spain, (pp. 1053–1056).
- Ménard, L., Perrier, P., & Aubin, J. (2016). Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. *The Journal of the Acoustical Society of America*, 139(5), 2514–2531.
- Messum, P., & Howard, I.S. (2012). Speech development: Toddlers don't mind getting it wrong. *Current Biology*, 22(5), R160–R161.

- Mitsuya, T., MacDonald, E.N., & Munhall, K.G. (2014). Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, 135(5), 2986–2994.
- Mitsuya, T., MacDonald, E.N., Munhall, K.G., & Purcell, D.W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, 138(1), 413–424.
- Mitsuya, T., MacDonald, E.N., Purcell, D.W., & Munhall, K.G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America*, 130(5), 2978–2986.
- Mitsuya, T., Munhall, K.G., & Purcell, D.W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America*, 141(4), 2758–2767.
- Mitsuya, T., & Purcell, D.W. (2016). Occlusion effect on compensatory formant production and voice amplitude in response to real-time perturbation. *The Journal of the Acoustical Society of America*, 140(6), 4017–4026.
- Mitsuya, T., Samson, F., Ménard, L., & Munhall, K.G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, 133(5), 2993–3003.
- Mollaei, F., Shiller, D.M., Baum, S.R., & Gracco, V.L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, 1646, 269–277.
- Mollaei, F., Shiller, D.M., & Gracco, V.L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, 28(12), 1668–1674.
- Munhall, K.G., MacDonald, E.N., Byrne, S.K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, 125(1), 384–390.
- Neufeld, C., Purcell, D., & Van Lieshout, P. (2013). Articulatory compensation to second formant perturbations. *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060097). ASA.
- Niziolek, C.A., & Guenther, F.H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, 33(29), 12090–12098.

- Osu, R., Hirai, S., Yoshioka, T., & Kawato, M. (2004). Random presentation enables subjects to adapt to two opposing forces on the hand. *Nature Neuroscience*, 7(2), 111–112.
- Palethorpe, S., Watson, C.I., & Barker, R. (2003). Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss. *The Journal of the Acoustical Society of America*, 114(2), 1055–1068.
- Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393.
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R.B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *Journal of Neuroscience*, 37(38), 9249–9258.
- Patri, J.-F., Perrier, P., Schwartz, J.-L., & Diard, J. (2018). What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. *PLoS Computational Biology*, 14(1), e1005942.
- Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4), 2338–2344.
- Perrier, P. (2012). Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control. In S. Fuchs, M. Weirich, D. Pape & P. Perrier (eds.). *Speech Planning and Dynamics*, Peter Lang Publishers, pp.191–238.
- Pfordresher, P.Q., & Palmer, C. (2006). Effects of hearing the past, present, or future during music performance. *Attention, Perception, & Psychophysics*, 68(3), 362–376.
- Pile, E.J.S., Dajani, H.R., Purcell, D.W., & Munhall, K.G. (2007). Talking under conditions of altered auditory feedback: does adaptation of one vowel generalize to other vowels. *Proceedings of the International Congress of Phonetic Sciences* (pp. 645–648).
- Purcell, D.W., & Munhall, K.G. (2008). Weighting of auditory feedback across the English vowel space. *Proceedings of the International Seminar on Speech Production* (Vol. 8, p. 389–392).

- Purcell, D.W., & Munhall, K.G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2), 966–977.
- Purcell, D.W., & Munhall, K.G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297.
- Reilly, K.J., & Dougherty, K.E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*, 134(2), 1314–1323.
- Reilly, K.J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology*, 118(5), 2925–2934.
- Rochet-Capellan, A., & Ostry, D.J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *Journal of Neuroscience*, 31(7), 2657–2662.
- Rochet-Capellan, A., Richer, L., & Ostry, D.J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717.
- Sakata, J.T., & Brainard, M.S. (2009). Social context rapidly modulates the influence of auditory feedback on avian vocal motor control. *Journal of Neurophysiology*, 102(4), 2485–2497.
- Sato, M., & Shiller, D.M. (2018). Auditory prediction during speaking and listening. *Brain and Language*, 187, 92–103.
- Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., & Gracco, V.L. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, 227(2), 275–288.
- Scheerer, N.E., Tumber, A.K., & Jones, J.A. (2015). Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *Journal of Neurophysiology*, 115(2), 826–832.
- Schuerman, W.L., Meyer, A.S., & McQueen, J.M. (2017). Mapping the speech code: Cortical responses linking the perception and production of vowels. *Frontiers in Human Neuroscience*, 11, 161.
- Schuerman, W.L., Nagarajan, S., & Houde, J. (2015). Changes in consonant perception driven by adaptation of vowel production to

- altered auditory feedback. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Schuerman, W.L., Nagarajan, S., McQueen, J.M., & Houde, J. (2017). Sensorimotor adaptation affects perceptual compensation for coarticulation. *The Journal of the Acoustical Society of America*, 141(4), 2693–2704.
- Sengupta, R., & Nasir, S.M. (2015). Redistribution of neural phase coherence reflects establishment of feedforward map in speech motor adaptation. *Journal of Neurophysiology*, 113(7), 2471–2479.
- Sengupta, R., & Nasir, S.M. (2016). The Predictive roles of neural oscillations in speech motor adaptability. *Journal of Neurophysiology*, 115(5), 2519–2528.
- Sengupta, R., Shah, S., Gore, K., Loucks, T., & Nasir, S.M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia*, 93, 242–250.
- Shadmehr, R., & Mussa-Ivaldi, F.A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, 14(5), 3208–3224.
- Shih, T., Suemitsu, A., & Akagi, M. (2011). Influences of transformed auditory feedback with first three formant frequencies. *International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11)*.
- Shiller, D.M., Lametti, D., & Ostry, D.J. (2013). Auditory plasticity and sensorimotor learning in speech production. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060150). ASA.
- Shiller, D.M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1308–1315.
- Shiller, D.M., Rvachew, S., & Brosseau-Lapr e, F. (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech-Language Pathology & Audiology*, 34(3), 181–192.
- Shiller, D.M., Sato, M., Gracco, V.L., & Baum, S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103–1113.

- Shum, M., Shiller, D.M., Baum, S.R., & Gracco, V.L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience*, *34*(11), 1817–1822.
- Smotherman, M., Zhang, S., & Metzner, W. (2003). A neural basis for auditory feedback control of vocal pitch. *Journal of Neuroscience*, *23*(4), 1464–1477.
- Sober, S.J., & Brainard, M.S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience*, *12*(7), 927–931.
- Stratton, G.M. (1897). Vision without inversion of the retinal image. *Psychological Review*, *4*(4), 341–360.
- Terband, H., & Van Brenk, F. (2015). Compensatory and adaptive responses to real-time formant shifts in adults and children. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Terband, H., Van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, *51*, 64–77.
- Thibeault, M., Ménard, L., Baum, S.R., Richard, G., & McFarland, D.H. (2011). Articulatory and acoustic adaptation to palatal perturbation. *The Journal of the Acoustical Society of America*, *129*(4), 2112–2120.
- Tourville, J.A., Cai, S., & Guenther, F.H. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060180). ASA.
- Tourville, J.A., Reilly, K.J., & Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*(3), 1429–1443.
- Tremblay, S., Houle, G., & Ostry, D.J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, *28*(10), 2426–2434.
- Tremblay, S., Shiller, D.M., & Ostry, D.J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869.
- Trudeau-Fisette, P., Tiede, M., & Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: Acoustic and articulatory data. *PloS one*, *12*(7), e0180300.

- van den Bunt, M.R., Groen, M.A., Ito, T., Francisco, A.A., Gracco, V.L., Pugh, K.R., & Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit. *Journal of Speech, Language, and Hearing Research*, 60(3), 654–667.
- van Vugt, F.T., & Ostry, D.J. (2018). The structure and acquisition of sensorimotor maps. *Journal of Cognitive Neuroscience*, (Early Access), 1–17.

- Vaughn, C., & Nasir, S.M. (2015). Precise feedback control underlies sensorimotor learning in speech. *Journal of Neurophysiology*, *113*(3), 950–955.
- Villacorta, V.M., Perkell, J.S., & Guenther, F.H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306–2319.
- Wei, K., Yan, X., Kong, G., Yin, C., Zhang, F., Wang, Q., & Kording, K.P. (2014). Computer use changes generalization of movement learning. *Current Biology*, *24*(1), 82–85.
- Wong, S.M., Domangue, R.J., Fels, S., & Ludlow, C.L. (2017). Evidence that an internal schema adapts swallowing to upper airway requirements. *The Journal of Physiology*, *595*(5), 1793–1814.
- Yates, A.J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*(3), 213–232.
- Zheng, Z.Z., Vicente-Grabovetsky, A., MacDonald, E.N., Munhall, K.G., Cusack, R., & Johnsrude, I.S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *Journal of Neuroscience*, *33*(10), 4339–4348.

3. Etude 1 - Transfert d'apprentissage et unités de parole chez l'adulte

Nous avons vu au travers de la revue de littérature du premier chapitre que différentes visions existent quant au niveau de granularité des unités assurant l'articulation de la parole chez l'adulte. Ces unités, que nous supposons sensorimotrices, sont-elles des mots, des syllabes, des phonèmes, ou encore des gestes articulatoires ?

Dans ce chapitre, nous apportons de nouveaux éléments de réponse à cette question, à travers une étude expérimentale chez l'adulte. L'étude repose sur le paradigme d'adaptation et de transfert auditorimoteur en réponse à une perturbation de formants, dont la revue de littérature a fait l'objet du 2^{ème} chapitre. C'est la seule étude publiée à ce jour, à notre connaissance, qui utilise ce paradigme pour contraster l'importance de différentes unités phonologiques, en particulier le mot, la syllabe et le phonème.

76 adultes francophones ont suivi un entraînement pour changer leur prononciation de la syllabe /be/ en réponse à une perturbation du retour auditif, soit en d'autres termes pour provoquer un apprentissage auditorimoteur. Nous avons ensuite évalué selon le groupe l'amplitude du transfert d'apprentissage de /be/ vers les bisyllabiques /bepe/, /pebe/ et /bebe/, ainsi que le post-effet sur /be/ prononcé de manière isolée. Cela a permis de contraster l'amplitude du transfert au niveau du mot, de la syllabe et du phonème, selon la position dans le mot. Nos analyses statistiques ont porté sur l'amplitude des phénomènes et leur évolution au fil des répétitions.

Les résultats montrent qu'il y a eu un transfert significatif de /be/ vers toutes les syllabes des mots de test. Ce transfert était plus grand vers /be/ que vers /pe/ en position initiale du mot, et aussi plus grand en position initiale que finale. La dynamique du transfert sur /be/ en première position dans un item bisyllabique était en outre différente du post-effet sur /be/.

Le transfert significatif sur une autre syllabe que la syllabe d'entraînement, mais contenant la même voyelle, suggère l'existence d'une représentation phonémique, ou infra-syllabique. Le constat que le transfert sur la syllabe d'entraînement est plus élevé que sur l'autre syllabe avec la même voyelle indique une contribution de l'unité syllabique dans l'articulation de la parole. La différence observée entre le post-effet et le transfert sur la même syllabe est le signe d'une influence du mot dans la production de parole. En conclusion, notre étude montre que les mots, les syllabes et les phonèmes jouent tous un rôle simultanément dans l'articulation de la parole. L'effet de position observé pose de nouvelles questions appelant d'autres études expérimentales.

Cette étude est présentée sous la forme d'un article de journal :

Caudrelier, T., Schwartz, J., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does it tell us about speech production units? *Journal of Speech Language and Hearing Research*, 1–13.

Deux actes de conférences internationales à comité de lecture et une communication dans une conférence internationale ont également été consacrés à une partie des résultats de la même étude expérimentale :

Caudrelier, T., Perrier, P., Schwartz, J. L., & Rochet-Capellan, A. (2016). Does auditory-motor learning of speech transfer from the CV syllable to the CVCV word? In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 08–12–Sept, pp. 2095–2099). <http://doi.org/10.21437/Interspeech.2016-262>

Caudrelier, T., Perrier, P., Schwartz, J., Savariaux, C., & Rochet-Capellan, A. (2016). De bé à bébé: le transfert d'apprentissage audiotri-moteur pour interroger l'unité de production de la parole. In *JEP TALN RECITAL 2016* (Vol. 1, pp. 101–109).

Caudrelier, T., Schwartz, J.-L., Perrier, P., Savariaux, C., & Rochet-Capellan, A. (2015). From Sensorimotor Experience To Speech Unit -Adaptation to altered auditory feedback in speech to assess transfer of learning in complex serial movements. In *SFN - Neuroscience, Oct 2015, Chicago, Unites States*. Retrieved from <https://hal.archives-ouvertes.fr/hal-01221491>

Research Article

Transfer of Learning: What Does It Tell Us About Speech Production Units?

Tiphaine Caudrelier,^a Jean-Luc Schwartz,^a Pascal Perrier,^a
Silvain Gerber,^a and Amélie Rochet-Capellan^a

Purpose: Words, syllables, and phonemes have each been regarded as basic encoding units of speech production in various psycholinguistic models. The present article investigates the role of each unit in the interface with speech articulation, using a paradigm from motor control research.

Method: Seventy-six native speakers of French were trained to change their production of /be/ in response to an auditory feedback perturbation (auditory–motor learning). We then assessed the magnitude of learning transfer from /be/ to the syllables in 2 pseudowords (/bepe/ and /pebe/) and 1 real word (/bebe/) as well as the aftereffect on the same utterance (/be/) with a between-subjects design. This made it possible to contrast the amplitude of transfer at the levels of the utterance, the syllable, and the phoneme, depending on

the position in the word. Linear mixed models allowed us to study the amplitude as well as the dynamics of the transfer and the aftereffect over trials.

Results: Transfer from the training utterance /be/ was observed for all vowels of the test utterances but was larger to the syllable /be/ than to the syllable /pe/ at word-initial position and larger to the 1st syllable than to the 2nd syllable in the utterance.

Conclusions: Our study suggests that words, syllables, and phonemes may all contribute to the definition of speech motor commands. In addition, the observation of a serial order effect raises new questions related to the connection between psycholinguistic models and speech motor control approaches.

Speaking requires the translation of a given message into a sequence of lexical units, which are then associated with a set of time-variant motor commands producing the appropriate sequence of sounds. A key step of this process is the decomposition of the lexical sequence into smaller constituents that map onto the motor system. This decomposition is called “phonological encoding” and has been a topic of interest and debates for years in psycholinguistics. A key issue in these debates is the nature of the units that interface the linguistic input with speech articulation, and two major models have been strongly influential in this field (Dell, 1986; Levelt, 1999).

Dell’s model holds that phonological encoding is performed by a hierarchically organized network connecting, from top to bottom, morphemes, syllables, syllabic constituents, phonemes, and features. In this model, the syllabic structure “onset–nucleus–coda” provides the frame of the

phonological encoding. The basic phonological units are phonemes associated with their position in the constituents of the syllables, namely, the onset, the nucleus, or the coda. The association of phonemes and their position within the syllable structure, described in terms of phonological features, are suggested to form the interface with speech articulation.

Levelt’s (1989) model, which also begins at the level of lexical concepts, attributes as well a key role to the syllable in phonological encoding. However, Levelt contests Dell’s proposal of a phonological encoding stage that would correspond to an “empty” syllabic skeleton filled by phonemes at a later stage. Levelt’s model rather suggests that the earliest level of phonological encoding is the specification of segments in the form of well-identified phonemes and that the syllabification process integrates which phonemes are the basic constituents of the syllable. This conceptual framework leads to the hypothesis that, for frequently used syllables, the articulatory gestures underlying the production of the syllable as a whole are stored as motor routines in a so-called “mental syllabary” (see also Levelt & Wheeldon, 1994). Thus, for the production of frequently used syllables, chunks corresponding to full syllabic motor schemas form the interface with articulation. For new or rarely used syllables, this interface could

^aUCNRS, Grenoble INP, GIPSA-lab, Institute of Engineering, University Grenoble Alpes, France

Correspondence to Tiphaine Caudrelier:
tiphaine.caudrelier@gipsa-lab.fr

Editor-in-Chief: Julie Liss

Editor: Bharath Chandrasekaran

Received April 23, 2017

Revision received October 13, 2017

Accepted March 22, 2018

https://doi.org/10.1044/2018_JSLHR-S-17-0130

Disclosure: The authors have declared that no competing interests existed at the time of publication.

possibly consist of phonemes that would be assembled online into syllables (Cholin, Levelt, & Schiller, 2006).

Conflicting empirical data contributed to raise a “syllable paradox” (Dell, 1986, p. 294; Levelt & Wheeldon, 1994, p. 241) with regard to the role of the syllable in phonological encoding, either as a generic skeleton filled by sub-syllabic units or as fully specified syllabic chunks. Several methods have been used to investigate this issue in psycholinguistics. A classical method is the analysis of speech error patterns. As a matter of fact, most errors are claimed to be exchanges of phonemes sharing the same position within the syllable structure (Shattuck-Hufnagel & Klatt, 1979), suggesting that both a syllable frame level and a phoneme content level operate in conjunction during phonological encoding (Nootheboom & Quené, 2015; Shattuck-Hufnagel, 1992). Processing time in speech production studies has also provided evidence for both the role of syllabic structure and syllabic content. Thus, the repetition speed of two pseudo-words is higher when they share the same syllabic structure in initial position (Sevold, Dell, & Cole, 1995), and masked priming studies have highlighted the influence of the generic syllabic structure independently of its phonemic content (Ferrand & Segui, 1998; Ferrand, Segui, & Grainger, 1996) as well as the role of the phonemic content in phonological encoding (Manoiloff, Segui, & Hallé, 2016). Besides, latency effects related to the frequency of linguistic units have also been instrumental in the debate related to the nature of speech units. The fact that high-frequency syllables correspond to lower production latencies (Cholin et al., 2006) has been interpreted as evidence for the existence of stored articulatory routines for such high-frequency syllables.

Beyond the “syllable paradox” and debates about syllables versus phonemes as basic units in phonological encoding, other studies in speech production have shed light on the possible role of larger units such as words. As a matter of fact, it appears that the sound changes affecting phonemes embedded in a word occur at different rates according to the frequency or grammatical status of the embedding word (Bybee, 2002). Thus, in American English, the deletion of the final stop consonants /t/ and /d/ occurs in verbs’ past participles at a faster rate when the verb is frequently used than when it is not, and it occurs at a slower rate in verbs’ past participles than in contracted negations, which are among the most frequently used grammatical expressions. These observations were interpreted by Bybee (2002) as evidence supporting the hypothesis that, in speech production, highly frequent words are not represented as a set of phonemes but rather as “exemplar clusters of their own” (p. 287). This hypothesis is consistent with those among the exemplar-based theories of speech that consider words as basic phonological units stored in the brain in the form of exemplars, that is, entities perceived or produced in a given situation, rather than generic prototypes (Välismaa-Blum, 2009). In this case, word exemplars would be at the interface between phonology and articulation.

In summary, psycholinguistic studies and speech production data, based on the analysis of speech error patterns, repetition and priming paradigms, and frequency and

grammatical effects, led to views that might appear contradictory about the nature of the basic phonological units that interface with speech articulation. Two possible interpretations are that various units could contribute in parallel to the definition of speech motor commands or that, depending on the method used, studies in psycholinguistics could actually look into representations that are not at the same level in the process of phonological encoding during speech production (see Cholin et al., 2006). Thus, speech errors, in particular, phoneme exchanges, may give information about units involved in serial order control, but not necessarily about the units ultimately tied to articulation. Priming studies could as well access to levels before the interface with speech motor commands.

Yet, understanding the nature of speech units enabling the translation of a sequence of lexical units into motor commands has key implications in the clinical field: It would help improving the treatment of various developmental speech disorders and, in particular, apraxia of speech (AOS). Treatment parameters usually include the selected linguistic unit for training (e.g., phonemes, words, or whole speech sequence), along with target selection (e.g., specific phonemes or words to practice pronouncing), to achieve the greatest therapeutic benefits (Crosbie, Holm, & Dodd, 2005). For example, training to articulate an isolated phoneme has often been used by speech therapists in the treatment of severe AOS. However, recent findings suggest that practicing the production of a single consonant does not lead to improvements in the production of this consonant embedded in various syllables produced by those patients (Aichert & Ziegler, 2013). In other words, this training does not transfer to other speech conditions. Choosing the syllable as the target unit seems more efficient because transfer of learning was observed from consonant–vowel (CV) syllables to words containing the specific training syllables.

The notion of transfer of learning is not only at the heart of rehabilitation methods for which recovery observed in speech therapists’ offices has to be associated with recovery in the daily life of patients. Transfer of sensorimotor learning from a speech unit to the same unit in a different word or to other units is also an experimental paradigm that could be used with typical speakers to directly address the nature of the representations involved in speech production. Transfer may occur when a common mental representation (e.g., the representation of a specific phoneme or syllable) is used to produce both the learning utterance and the transfer utterance. Hence, according to Houde and Jordan (1998), transfer of sensorimotor learning from a given vowel (e.g., /ε/) in a consonant–vowel–consonant (CVC) context (e.g., “pep”) to the same vowel in another context (e.g., “set” or “forget”) is a way to directly assess “the putative hierarchical, segmental control of speech production” (Houde & Jordan, 1998, p.1214). In their seminal study, Houde and Jordan (1998) induced sensorimotor learning in typical speakers using an auditory feedback perturbation paradigm. Subjects were asked to repeat CVC words (e.g., “pep”) while their auditory feedback was altered in real time so that the vowel /ε/ sounded more like /a/ (e.g., changing

“pep” to “pap”). After many repetitions, a large majority of subjects learned to partially compensate for the perturbation by pronouncing something closer to “pip,” which suggests a sensorimotor recalibration designated as “adaptation.” After the perturbation is stopped, a change in the production of the training utterance can persist. This change, referred to as the “aftereffect,” demonstrates that the procedure triggers learning. The auditory–motor learning paradigm has been extensively used to study speech motor control (e.g., Max & Maffett, 2015; Purcell & Munhall, 2006). Houde and Jordan also showed a partial transfer of auditory learning from /ɛ/ in a set of training words to the same vowel in other CVC words, suggesting that the vowel could be a speech production unit (Houde & Jordan, 1998, 2002). Similar observations were made later both in English and in Mandarin (Cai, Ghosh, Guenther, & Perkell, 2010; Villacorta, Perkell, & Guenther, 2007).

Transfer hence appears as an experimental tool likely to directly address the relationship between the phonological units involved in phonological encoding and the associated articulatory commands. However, to our knowledge, previous studies of sensorimotor learning transfer involved a single phonological level, and no study has been designed with the specific intent of contrasting the role of different speech units in the specification of speech motor commands. Therefore, we extended the approach introduced by Houde and Jordan (1998) to contrast transfer effects at three levels: global utterances (larger than syllables), CV syllables (regarded as chunks that would be coded globally), and phonemes (which would be coded individually in a given syllabic CV structure). Our experimental protocol consisted of assessing sensorimotor transfer from a given vowel within a given high-frequency CV syllable to disyllabic CVCV utterances containing this syllable in the first or second position, together with another syllable composed of the same vowel and a different consonant. This protocol enables us to investigate whether the decomposition that is tied to speech articulation is composed of (a) subsyllabic units (here, phonemes) in a given position in the CV syllabic frame, as suggested by Dell’s (1986) model; (b) syllables encoded as chunks, in line with Levelt’s proposal for high-frequency syllables (Levelt, 1989); or (c) other units larger than syllables (words or utterances for instance), as argued by some exemplar theories.

Method

Participants

Seventy-six native speakers of French (33 women and 43 men, age range = 18–35 years) took part in the experiment. Subjects had no reported speech, language, or hearing disorders and were naive to the purpose of the experiment. French from metropolitan France was their mother tongue, and none of them learned or was extensively exposed to a second language in early childhood (before 6 years old). Rather than using geographical origin within metropolitan France as a selection criterion, because accents

are variable within regions, we explicitly tested the contrast /e/–/ɛ/ in production, as described in the procedure. Two channels were used to find participants: flyer distribution on the campus of University Grenoble Alpes, France, and an e-mail sent via the diffusion list of RISC (Relais d’Information sur les Sciences de la Cognition), which is a research unit affiliated with the CNRS (Centre National de la Recherche Scientifique [French National Research Institute]). Anybody can subscribe to this diffusion list on the website of RISC. As a result, participants were students or young professionals, and all of them had a baccalauréat degree (French high school diploma). Socioeconomic status was recorded, but it was not a selection criterion because it has not been shown to influence auditory–motor adaptation or transfer, to our knowledge. Participants were split into four groups according to the utterance used to evaluate the transfer (see below). They were randomly assigned to one of the groups, with the female-to-male ratio held consistent among the groups.

Equipment and Task

The participants were seated in front of a monitor in a soundproof room and wore headphones with a built-in microphone (Sennheiser HME 26-II-600). Utterances were displayed on a monitor. The speakers were instructed to read them aloud in a natural way, without shouting or whispering. They heard their own voice mixed with speech-shaped noise in the headphones.

The first two formants (F1 and F2) in speakers’ productions were shifted in real time (with a nondisruptive 14-ms delay [Yates, 1963]) using the Audapter system (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008). This auditory feedback perturbation resulted in shifting the vowel /e/ toward /ɛ/ by increasing F1 by 27% and decreasing F2 by 10%.

Experimental Design and Hypotheses

Training Utterance

The training utterance was the monosyllabic logatome /be/ in all groups (written “bé”). The vowel /e/ was selected mainly because of auditory perturbation constraints: The vowel had to be roughly in the middle of two existing French vowels in the (F1, F2) space, and back vowels were rejected because F1 and F2 are harder to detect for such vowels. The consonant /b/ was chosen to limit consonant-to-vowel coarticulation (Recasens, 1997). Several other linguistic factors were also taken into consideration. First, the syllable /be/ is as frequent at the beginning of French words as it is in the middle or at the end, as verified using the Lexique 2 database (New, Pallier, Brysbaert, & Ferrand, 2004). Second, it has been suggested that the magnitude of the adaptation to auditory perturbations is influenced by the frequency of auditory and articulatory neighbors (i.e., the word resulting from the auditory feedback alteration, here /be/, and the word that has to be produced to compensate for it, here /bi/, respectively; Frank, 2010). Accordingly, the

training word /be/ was chosen so that its neighbors /be/ and /bi/ were of low frequency, thus limiting their influence on adaptation.

Test Utterances

As shown in Table 1, the utterances used to evaluate the transfer of sensorimotor learning (“test utterances” henceforth) differed according to the group and were chosen to contrast the transfer at the level of the phoneme, the CV syllable, and the word. In a first group, it was the same as the training utterance /be/ (this group will be referred to as the *be_transfer* group below). The pseudowords /bepe/ and /pebe/ (written “bépé” and “pébé”) were used in two other groups (the *bepe_transfer* group and the *pebe_transfer* group, respectively). These pseudowords allowed us to study the transfer to the same syllable in a different utterance (e.g., from /be/ to /bepe/) and the transfer to the same vowel in another CV syllable (e.g., from /be/ to /pe/), while changing the position of the syllable /be/ and /pe/ within the utterance. /p/ is a bilabial stop consonant that is articulatorily very close to /b/ and differs mainly by voice onset time. The French word /bebe/ (“bébé,” which means “baby”) was chosen to evaluate the influence of lexical status (real word vs. pseudoword) on the transfer in the last group referred to as the *bebe_transfer* group. As a matter of fact, lexical status has been shown to influence the magnitude of auditory–motor adaptation (Bourguignon, Baum, & Shiller, 2014). Besides, /pe/ was not selected as a test utterance because it would not provide any additional information about the research question.

Predictions

The experimental design makes it possible to contrast the transfer magnitude at different levels: same utterance as in training, same syllable in a different utterance, and same phoneme in a different CV syllable. The effect of the position in the utterance together with the lexical status can also be tested.

Our main predictions are that, (a) if the interface between phonological encoding and speech articulation is at the level of phonemes (or any unit smaller than syllables, e.g., phonetic features), then a transfer from the syllable /be/ to the syllable /pe/ should occur; (b) if it is at the level

of syllable chunks, there will be more transfer to the syllable /be/ than to the syllable /pe/; and (c) if it is at the level of words/utterances, the retention of learning in the utterance /be/ will be greater in magnitude than the transfer to any syllable in the other test utterances (/bepe/, /pebe/, and /bebe/). In addition, we have two complementary questions. The position of the syllable within a larger utterance may have an effect on transfer, although we have no specific assumption about the precise nature of this effect. Finally, in line with the observations that the perceptual bias toward real words increases tolerance to auditory perturbation (Bourguignon et al., 2014; Ganong, 1980), we expect transfer to be stronger in the real word /bebe/ than in the pseudowords /bepe/ and /pebe/.

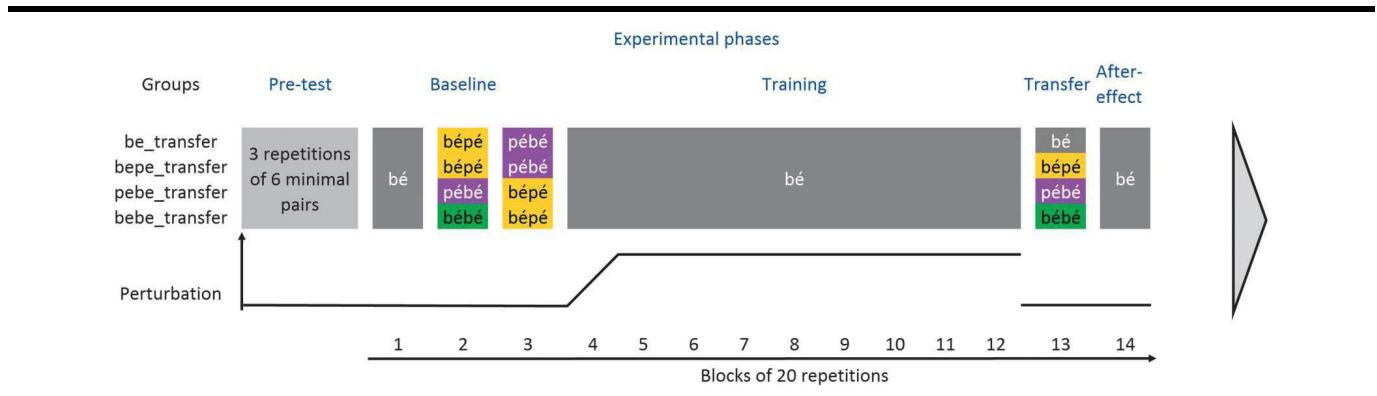
Experimental Procedure

The experimental procedure is described in Figure 1. In a pretest (Blocks 1–3), speakers pronounced six minimal pairs (e.g., real French words “épée” [“sword”] and “épais” [“thick”], pronounced /epe/ and /epe/) to explore the contrast /e/-/ɛ/ in each speaker. This was done to be able to control whether it had any impact on the results of the study, because the strength of this contrast is quite variable among French speakers (but so far, no study has shown any impact of a phonemic contrast over the adaptation or transfer of auditory–motor learning, to our knowledge).

The experiment then consisted of 14 blocks of 20 trials each. At the end of each block, the participants had to press a key to continue the experiment. In the baseline phase (Blocks 1–3), all groups produced three different utterances, including the training utterance (/be/) and their test utterance with unaltered feedback. This baseline allows us to measure speakers’ production before learning, providing a reference against which we could assess transfer and aftereffect magnitudes. During the training phase, the auditory feedback alteration was gradually set up (Block 4) and then held at its maximal magnitude (Blocks 5–12) while all groups were repeating /be/. The perturbation was then stopped, and speakers had to produce the test utterance assigned to their group (transfer phase, Block 13). Finally, all groups pronounced /be/ again to test for the aftereffect (i.e., the trace of the auditory–motor learning

Table 1. Experimental design and factors tested.

Factor	Group							
	be_transfer		bepe_transfer		pebe_transfer		bebe_transfer	
Test word utterance	bé (/be/)		bépé (/bepe/)		pébé (/pebe/)		bébé (/bebe/)	
Utterance	= training		≠ from training utterance		≠ from training utterance		≠ from training utterance	
Lexical status	Pseudo-word		Pseudo-word		Pseudo-word		Real word	
Syllable	/be/	/be/	/be/	/pe/	/pe/	/be/	/be/	/be/
	(= training)	(= training)	(= training)	(≠ syllable)	(= training)	(= training)	(= training)	(= training)
Position in the word	1	1	1	2	1	2	1	2
Phoneme	/e/	/e/	/e/	/e/	/e/	/e/	/e/	/e/
	(= training)	(= training)	(= training)	(= training)	(= training)	(= training)	(= training)	(= training)

Figure 1. The experiment consisted of five phases, with each number corresponding to one block of 20 trials.

and its decline, Block 14). It should be noted that, in the *be_transfer* group, the utterance is the same in both the transfer and aftereffect blocks.

Measurements

F1 and F2 were assessed with Praat software (Boersma, 1993) using linear predictive coding computed on a sliding window of 25 ms with a 1-ms time step. The values were averaged over 30 consecutive points in the stable part of each vowel production (Rochet-Capellan & Ostry, 2011).

The duration of the voiced part of the syllable, as well as the intensity and pitch in the stable part of each vowel, was also measured using Praat to control for any effect of prosody on adaptation and transfer in relation to syllable position within words or nonwords.

All vowel utterances produced in the various experimental phases were characterized by a global index (F2–F1), with F1 and F2 expressed in Barks, a usual auditory measurement of frequency that represents human perceptual scale better than Hertz (Schroeder, Atal, & Hall, 1979; see Equation 1).

$$f_{\text{Bark}} = 7 * \operatorname{argsinh}\left(\frac{f_{\text{Hertz}}}{650}\right) \quad (1)$$

This F2–F1 index characterizes the position of a vowel on the open–close dimension for front vowels (Fant, 1960). Then, the main measure chosen to assess transfer magnitude was the percentage of F2–F1 change compared with the corresponding F2–F1 value measured in the baseline phase for the same utterance in the same speaker. For instance, the first /e/ of each repetition of /bebe/ in the transfer phase was expressed relative to the average first /e/ of /bebe/ productions in the baseline. This percentage of change will be referred to as F2–F1 change in the Results section.

Statistical Analysis: Linear Mixed Models

First, the adaptation in each speaker was estimated by comparing F2–F1 in the last 40 trials of the training

phase with the baseline using a one-tailed paired *t* test. This allowed us to check that adaptation occurred in a given speaker. A proportion equality test (using the `prop.test` function of R software) was run to test whether the proportion of adapted speakers was equivalent in each group. The analyses carried out using R software (Version 3.2.0; R Core Team, 2016) to evaluate adaptation, transfer, and aftereffects were as follows.

General Approach With Linear Mixed Models

Because we wanted to assess the evolution of F2–F1 over trials from the adaptation to the transfer and after-effect blocks, we built several linear mixed models (LMMs) using the `lme` function from the `nlme` package of R (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2007). LMMs enabled us to take the temporal structure of our data into account (Liu, Cripe, & Kim, 2009). In all but one case (specified in the main effects section), we followed a backward deletion approach (Mundry & Nunn, 2009) to select the best LMM using likelihood ratio test. In all LMMs, F2–F1 change was the variable to explain, and the speaker was considered as a random factor.¹ Post hoc tests were adjusted for multiple comparisons, with the `glht` function from the `multcomp` package of R (Hothorn, Bretz, & Westfall, 2008). In all LMMs, the trial was considered as a numerical factor (ranging from 0 to 19). This factor was used to evaluate the evolution over the block. Evolution was then characterized by a slope through linear regression, assuming a linear variation of F2–F1 within the block.

Main Effects

To evaluate the adaptation effects, the considered factors were the group and the trial. The transfer effects

¹As there were repeated measures over time for a given participant, residuals structure could depend on the trial. To check the absence of autocorrelation in residuals, autocorrelograms were plotted. When necessary, the residual variance–covariance structure was modeled to allow within-participant correlated errors.

required two separate analyses because the experimental design was not fully balanced (e.g., there is neither training with /pe/ nor transfer toward a single /pe/). In a first analysis restricted to the *bepe_transfer*, *pebe_transfer*, and *bebe_transfer* groups, we tested the influence of the word, the syllable, and the position. This was done by comparing the two best models derived from Syllable \times Position \times Trial² versus Word \times Position \times Trial using the Akaike information criterion that is suitable for nonnested models. In a second analysis, /be/ in the *be_transfer* group was compared with /be/ produced in the first position in the other groups.

To correct for multiple comparisons in an effective way, the two best models derived from these analyses were merged by adding them. This merge did not affect the regression coefficients of the two models (which were independent). Then, we performed post hoc tests.

In the aftereffect block, the considered factors were the group and the trial. An LMM was built to test in each group whether the production of /be/ in the aftereffect block followed the same trend as the production of the syllable /be/ in the transfer block.

Additional Analyses

First, it is important to note that all the analyses performed on F2–F1 were also performed on F1–F0 (F1 and F0 also expressed in Bark), another possible index of the open–close dimension, possibly more closely related to mandible aperture because F2–F1 also involves tongue position (Traunmüller, 1981). However, the analyses with F1–F0 provided essentially the same results and will not be reported in the Results section.

In addition, prosodic cues, namely, vowel duration, pitch, and intensity, were analyzed to assess their potential role in the transfer phase, using three separate LMMs. The considered fixed factors were the test utterance and the syllable position. Then, possible correlations across participants between the intersyllabic difference (first vs. second syllable of each transfer word) in these prosodic cues and the intersyllabic difference in the transfer magnitude were investigated using Spearman's rank correlation.

Moreover, correlations across speakers between their /e/–/ɛ/ contrast evaluated from the pretest phase and the magnitude of their adaptation were assessed using Spearman's rank correlation. Contrast in the pretest was obtained by measuring the formant values for /e/ and /ɛ/ in the six uttered minimal pairs and computing the Euclidian distance between mean values for /e/ and /ɛ/ in the (F1, F2) space. Adaptation was expressed here as the Euclidian distance between the mean values for /e/ in the baseline and at the end of the training phase.

²Syllable \times Position \times Trial represents the models that include the three mentioned factors and the simple and double interactions between them.

Results

The evolution of F2–F1 by group over the last block of the training (adaptation), transfer, and aftereffect phases is shown in Figure 2. Figure 3 provides the parameters of the LMM in the transfer block.

Adaptation

We were expecting participants to compensate for the perturbation of their auditory feedback by increasing F2–F1 in /be/ over the course of the training phase. Because this study is focused on transfer of adaptation, only participants who showed a significant adaptation over 3% were selected in the analysis. Fifty-six speakers satisfied this constraint (~74%): 12 in the *be_transfer* group, 15 in the *bepe_transfer* group, 15 in the *pebe_transfer* group, and 14 in the *bebe_transfer* group. The proportion of adapted speakers was statistically equivalent in each group, $\chi^2(3) = 1.25, p = .74$.

At the end of the training phase, the average adaptation was 7.8% (mean \pm 0.8% (SD)). This adaptation was homogeneous among groups and stabilized (Figure 2, left) because the obtained LMM shows no significant effect of the group and no significant change over trials.

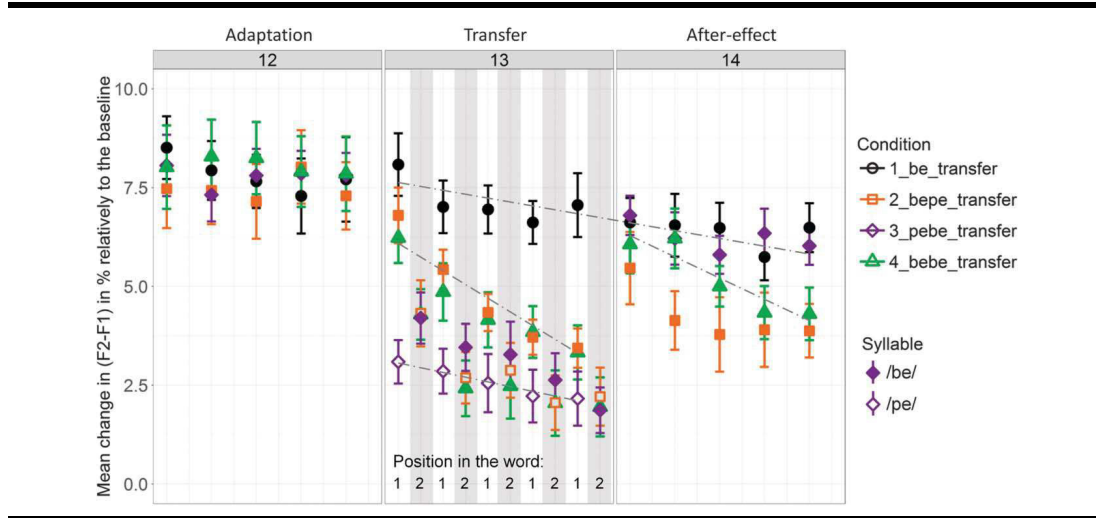
Transfer

Figure 2 shows the following trends in the transfer phase. First, some transfer occurred in all groups as evidenced by error bars being all well above the 0 line. The transfer to the monosyllable /be/ (*be_transfer* group, black markers) is much larger than the transfer that occurred in all other groups. It results in a gentle negative slope from the adaptation phase (first dotted line on the top of the figure). In the other groups, the transfer seems higher in the syllable /be/ in initial position (e.g., in /bepe/ or /bebe/) than in /be/ in the second position (in /pebe/ or /bebe/) or in /pe/ (in /bepe/ or /pebe/). It is displayed on Figure 2 by a lower dotted line starting more or less from the same position as the one in the *be_transfer* group, but with a steeper negative slope. The smallest transfer is observed in the syllable /pe/ in the first position (in /pebe/). This trend is suggested by the lowest dotted line in the figure.

The best model fitting transfer data on disyllabic words actually show an effect of the produced syllable (/be/ vs. /pe/), of its position in the word (first vs. second), and of their interaction on both the intercept and the decay slope. It also highlights the influence of the testing word being identical to the training word (largest transfer in monosyllabic /be/).

Post hoc comparisons reveal that, (a) the intercept (Figure 3a) of the transfer in /be/ was significantly higher than in /pe/ when produced in the first position in the word ($6.6\% \pm 0.9\%$ vs. $3.3\% \pm 1.2\%$; $z = 4.33, p < .001$); (b) the intercept in /be/ was also significantly higher in the first position than in the second position in the word ($6.6\% \pm 0.9\%$ vs. $4.3\% \pm 0.9\%$; $z = 3.72, p = .002$); (c) the decay (Figure 3b) was significantly faster (larger negative slope) in /be/ than in /pe/ when produced in the first position

Figure 2. Evolution of F2–F1 (in percent change in reference to baseline) over the last three blocks of the experiment (corresponding to the end of adaptation, transfer, and aftereffect phases), by group and by syllable. Each point represents the average value for a group and a syllable on four trials (with hence 5 points per syllable and per block in a given phase). In the transfer block, the second syllable of the utterance is shifted to the right to separate it from the first syllable, as indicated on the figure (“Position 1 vs. 2”). The dotted gray lines highlight the main trends followed by group–syllable–position combinations.



($-0.18\% \pm 0.04\%$ vs. $-0.07\% \pm 0.04\%$; $z = -2.793$, $p = .045$); and (d) the decrease was significantly slower in the test word /be/ than in the first syllable of both disyllabic words /bebe/ and /bepe/ ($-0.06\% \pm 0.07\%$ vs. $-0.18\% \pm 0.04\%$; $z = -2.778$, $p = .047$). No other significant effects were found.

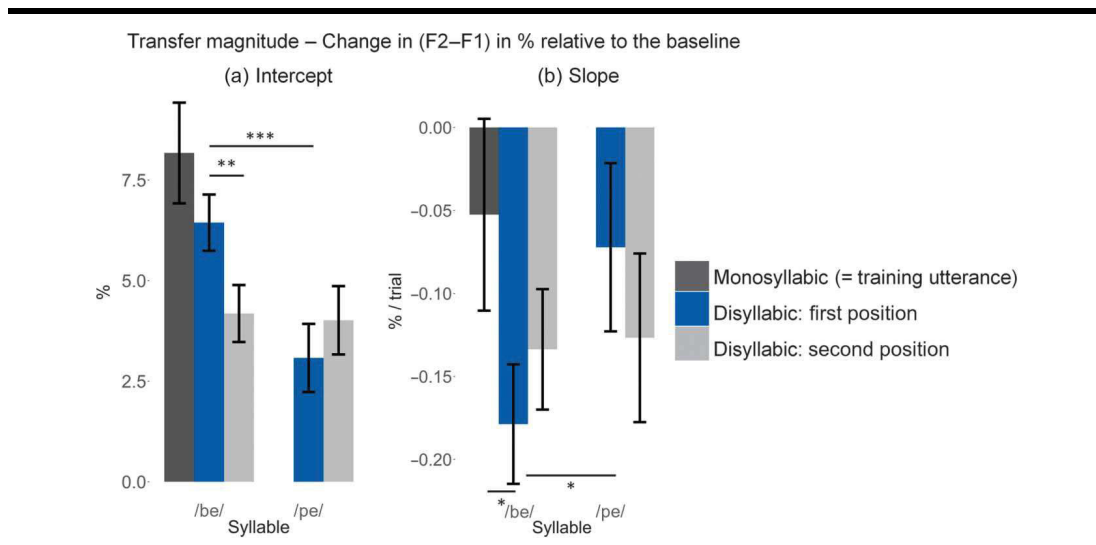
Aftereffect

Figure 2 shows that the aftereffects in the be_transfer and pebe_transfer groups seem to be in continuity with the

slow decrease of adaptation along the transfer phase for the monosyllabic /be/ (see the top dotted line on transfer and aftereffect). The aftereffect for the two other groups appears lower with a faster decay.

The best model fitting aftereffect data confirms an effect of the group on the evolution over the block. Post hoc comparisons showed no difference between groups in the intercept. However, the decay in the bebe_transfer group was significantly faster than in the be_transfer group ($-0.13\% \pm 0.05\%$ vs. $-0.02\% \pm 0.06\%$; $z = 2.88$, $p = .040$).

Figure 3. Parameter estimates of the linear mixed model in the transfer phase: (a) intercept and (b) slope. Results are given by syllable (/be/ vs. /pe/), utterance (monosyllabic vs. disyllabic), and position (first vs. second syllable in the disyllabic utterances). The bars represent the confidence interval of the estimates. The significance of post hoc comparisons is represented with asterisks (* $p < .05$, ** $p < .01$, *** $p < .001$).



The aftereffect in the *bepe_transfer* group seems to be different than in the *be_transfer* and *pebe_transfer* groups, but this difference was not statistically significant with regard to either the intercept or the slope.

The LMM associating transfer and aftereffect data by group showed the continuity in the evolution of F2–F1 change in the *be_transfer* group over the course of these two blocks ($p > .05$). In the other groups, a discontinuity was shown between the transfer and aftereffect blocks ($p < .005$ in each of these groups), whereas the decay speed was not different between blocks ($p > .05$ in all groups, including the *be_transfer* group).

Prosodic Cues and Transfer

The second syllable lasted significantly longer than the first syllable in all disyllabic test utterances (+30 ms; $z = 10.1$, $p < .001$). The intensity was higher in the first syllable than in the second one (-1.3 ; $z = -5.85$, $p < .001$). The pitch was significantly higher in the first than the second syllable, only in */bepe/* (+0.07 Barks; $z = 2.72$, $p < .01$) and */pebe/* (+0.17 Barks; $z = 6.68$, $p < .001$) but not in */bebe/* (+0.04 Barks; $z = 1.40$, $p = .16$). The observed prosodic pattern was consistent with the combination in French of a primary stress on the last syllable of a phrase, here associated with a longer duration, and an initial stress on the first syllables, associated with a pitch increase (see Jun & Fougeron, 2002, for more details about French prosody patterns). In addition, no correlation was found between the difference in transfer magnitude of the first versus second syllable and the intersyllabic difference in any of these cues in the transfer block (duration: $\rho = -0.02$, $p = .8$; pitch: $\rho = 0.03$, $p = .8$; intensity: $\rho = -0.09$, $p = .4$).

Contrast /e/–/ɛ/ in Production and Adaptation

No correlation was found across speakers between their */e/–/ɛ/* contrast in production (as assessed in the pretest) and the magnitude of their adaptation ($\rho = 0.19$, $p = .18$).

Discussion

Psycholinguistic models of speech production have proposed conflicting views about the units involved in phonological encoding and providing the interface with speech articulation. We argued that the apparent contradictions between empirical data might be due to the variety of paradigms used to address this question and the fact that most methods do not specifically focus on the interface with speech articulation. In line with Houde and Jordan (1998), this study investigated the role of various linguistic units in the specification of speech motor commands, by studying their influence on the transfer of auditory–motor learning. It was designed to assess whether motor commands are defined (a) at a subsyllabic level, which could be phonemes in a given syllabic position, in which case transfer would be observed at the level of the vowel (Hypothesis 1); (b) at

the level of syllables, represented as chunks, in which case the transfer magnitude should depend on the syllable (Hypothesis 2); and/or (3) at the level of words or, more generally, utterances, in which case transfer should depend on the word or the utterance (Hypothesis 3). The results tend to show that transfer of learning occurs at the three levels in parallel. In addition, transfer does not seem to depend strongly on the lexical status of the utterance but is highly influenced by the syllable position in a disyllabic utterance. These findings are discussed in the following sections.

What Are the Main Units of Speech Production in Adults?

First, significant transfer occurred in all groups, in all produced vowels regardless of the embedding syllable. Transfer therefore involves a lower level than the syllable. This level may be the phoneme—here, the vowel /e/ that constitutes the nucleus of CV syllables /be/ and /pe/. Importantly, it could also be a phonetic feature, such as the position of the tongue when pronouncing the vowel /e/, or any characteristic common to the syllables /be/ and /pe/ (e.g., all syllables with a CV structure or all syllables with a stop consonant). This study was not designed to disentangle these possibilities because it contrasted the magnitude of transfer neither across several consonantal and vowel features nor across various syllabic structures. To question the role of a syllabic frame as suggested by Dell's (1986) model, in relation to speech articulation, transfer from a phoneme in a given syllabic position (an onset consonant, for instance) to the same phoneme in another position (e.g., in coda) should be studied. However, the important point here is that subsyllabic components at the lowest level of the hierarchy of units considered in this study do intervene in transfer patterns. This result is in agreement with previous work showing transfer at the phoneme level (e.g., Houde & Jordan, 1998, 2002; see the introduction). Significant transfer has also been observed from the vowel /ɛ/ not only toward other front vowels such as /i/ or /a/ (Houde & Jordan, 1998) but also toward back vowels (Villacorta et al., 2007), suggesting that broader changes could also occur in auditory–motor learning. This is confirmed in a recent study showing that transfer may occur from one vowel to all the adjacent vowels within a certain range in the phonetic space (see Reilly & Pettibone, 2017).

At a higher level in the hierarchy, there was an effect of the syllables (/be/ vs. /pe/) in the magnitude of the transfer. Transfer was lower in /pe/ than in /be/, although the consonants /b/ and /p/ are quite similar from an articulatory standpoint, both of them being bilabial stop consonants, differing mainly in voice onset time. This result suggests that some transfer specifically occurs at the syllable level and therefore that syllable chunks are involved in the definition of motor commands.

Going one step further, some transfer seemed to occur at the level of a higher-level unit, here the word or the utterance. This claim is supported by conjointly analyzing both the transfer effect and the aftereffect. The decay of

the transfer in the syllable /be/ of /bebe/ was significantly faster than the decay observed in the monosyllabic /be/ in the transfer block. There was, however, no difference at the beginning of the block (as shown by the equal intercepts in both groups), but only in the decay (i.e., the slope). One could argue that the faster decay of transfer seen in /bebe/ compared with /be/ is driven by the fact that 20 repetitions of /bebe/ actually involved 40 productions of /be/, twice as many as the repetitions of the monosyllabic /be/. This may have led to a faster decrease of the effect magnitude over the transfer phase. If this was the case, there should be a difference in the intercept in the aftereffect block: After repeating /bebe/ in the transfer phase, speakers should show less aftereffect on /be/ than after repeating the monosyllabic /be/. Our results show no difference between the groups in the intercept of the aftereffect block. Rather, a break was observed between the production of /bebe/ during the transfer phase and the production of the utterance /be/ in the aftereffect: The changes in /bebe/ induced by the adaptation phase progressively decrease during the transfer phase, but this does not affect the aftereffect on the training utterance /be/. Overall, these observations support the hypothesis of a specific role of the word /bebe/ in transfer magnitude, and then, more generally of the existence of units larger than syllables interfacing with speech articulation. It should be noted that these observations were based on the statistical analysis of the evolution over the transfer and aftereffect phases. Previous works that have assessed the transfer of auditory–motor learning in an independent block, after the adaptation phase, have focused only on the first trials of the block (Rochet-Capellan, Richer, & Ostry, 2012). Our approach shows that the transfer dynamics itself and the transition to the aftereffect bring important insights.

The observed pattern of transfer could also be interpreted in terms of gradient of similarity between the training utterance and the test utterance. In English, the transfer at a vowel level from one CVC utterance to another was observed to be proportional to the acoustic similarity between the vowels in these CVC as measured before learning (Rochet-Capellan et al., 2012). However, in our experiment, no correlation was observed between (a) the transfer and (b) the acoustic distance in the F1–F2 space between the vowel in the training utterance and the vowel in the test utterance in the baseline. Other levels of similarity could potentially explain the patterns observed here, such as an interaction of both categorical variables (e.g., linguistic representations) and more continuous ones (e.g., articulatory or acoustic proximity). Further experimental research and also computational models should help in the understanding of the role of these different variables on transfer in sensorimotor learning.

Overall, the results suggest that the transfer of sensorimotor learning does not occur at a specific level in terms of the linguistic unit but rather at several levels in parallel. The interactions between these levels are quite complex. Thus, there may be several units that contribute to the definition of motor commands. This observation in adults is consistent with Vihman and Croft's developmental theory

(Vihman & Croft, 2007), according to which several speech production units may develop gradually and therefore coexist in adults (see also Tilsen, 2016). These results can also be related to the increasingly acknowledged view in speech perception research that various sublexical and lexical levels coexist and collaborate to ensure the adequate processing of speech inputs in the comprehension of both oral and written language (Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2016).

Real Word vs. Pseudoword

Our corpus was also designed to test the influence of the lexical status (real word vs. pseudoword) of the transfer word. The results do not show any difference in the intercept of the magnitude of transfer between the first syllable of /bebe/ and /bepe/ or between the second syllable of /bebe/ and /pebe/. However, the lexical status has been shown in the literature to influence the magnitude of auditory–motor adaptation: If the production of a real word is altered toward a sound close to a pseudoword, the magnitude of the adaptation is supposed to be lower than in the opposite situation (Bourguignon et al., 2014). Our data do not show any evidence of this difference in the transfer phenomena we have observed.

The Effect of the Position of the Syllable in the Word

The magnitude of the transfer to the syllable /be/ is significantly higher when /be/ is in the first position of the word than when it is in the second position. The large magnitude of this effect was not expected, and our study, to our knowledge, is the first one to display an effect of serial order on the magnitude of the transfer of speech auditory–motor learning. Several explanations could account for this effect—referred to as the position effect.

First, the position effect could be prosody driven. A stressed syllable could be produced in a more precise way than an unstressed syllable. In most participants, the second syllable lasted longer than the first one, which is consistent with French stress patterns (Jun & Fougeron, 2002). However, no correlation across speakers was observed between the stress pattern and the difference of transfer magnitude associated with the position of the syllable of interest in the transfer word. Therefore, a prosodic interpretation based on stress patterns cannot explain the differences in transfer magnitude driven by the syllable position in the word.

Alternatively, the position effect could be the consequence of an initialization effect. Many studies in speech motor control have emphasized the specific role played by the initial syllable in the word (MacNeilage & Davis, 2000). Movement initiation is generally regarded as more difficult than movement continuation and is perhaps performed by a specific subsystem within the motor system. Thus, the first syllable may be produced more carefully than the following ones, with a better goal specification, and therefore may preferentially carry the modification

triggered by auditory–motor learning, although the mechanism remains unclear. This could then explain the observation that the transfer to /be/ as the first syllable decays faster than to /be/ as a monosyllabic word. At the beginning of the transfer block, the initial syllable would be produced as a separate single syllable, requiring a specific initiation mechanism. Gradually, over several repetitions, initiation would become less difficult and the word would be produced as a whole.

The position effect may also be a serial effect that depends on auditory perturbation. The training was realized on a monosyllabic word. The perturbed syllable was thus the first one (and the only one) of the utterance. In this context, the speakers may have learned to modify more specifically the first syllable of the sequence. This is consistent with the view that phonological encoding is performed serially and incrementally, as supported by several priming experiments in speech production (Meyer, 1990; Roelofs, 1999).

Finally, a lower degree of transfer to the second syllable than to the first one could be driven by online speech control mechanisms based on auditory feedback (Hawco & Jones, 2010; Tourville, Reilly, & Guenther, 2008). While speaking, a prediction of the auditory outcome based on motor commands is believed to be compared with the acoustic realization. If a difference between the prediction and the acoustic realization is noticed for the first syllable, then online control mechanisms would tend to close this gap and to decrease the transfer to the next syllable.

To make an informed decision between these possible explanations of the observed position effect, further investigation is required. The paradigm of auditory–motor learning transfer could be applied to training words consisting of several syllables. Ideally, the auditory perturbation would be applied selectively to a syllable in a specific position in the word (e.g., the second one). The transfer to the same syllable in the first versus second position of another word would then be evaluated. So far, to our knowledge, the auditory–motor learning paradigm has not been implemented in humans with polysyllabic words. Yet, a similar paradigm with a pitch perturbation on a specific syllable in a given position of bird vocalizations was previously applied to zebra finches (Hoffmann & Sober, 2014). In humans, selective perturbation in terms of position in the sequence would raise new perspectives in studying how speech units are put into sequence from both phonological encoding and motor control perspectives.

Limitations and Perspectives

The present work provides evidence in favor of a convergence of subsyllabic, syllabic, and suprasyllabic factors in auditory–motor learning transfer, which we have related in the previous sections of this discussion to phonological units likely to be involved in the definition of motor commands in speech production. We have already mentioned in these sections some limitations and perspectives for experimental clarification, in relation, for instance, with the nature of subsyllabic units or the effect of syllable position in the word.

More broadly, some methodological limitations intrinsically associated to the auditory–motor learning transfer paradigm could have influenced the results. First, it is important to remember that transfer cannot be studied if there is no adaptation. Hence, all our data and conclusions are, by nature, restricted to subjects actually displaying significant adaptation. The adaptation rate observed in this study is 74%, because 56 of 76 speakers adapted to the perturbation by significantly increasing F2–F1. Although this adaptation rate does not seem very high, it is very similar to what has been observed in previous studies (Cai et al., 2010; Houde & Jordan, 2002). A possible reason some speakers do not adapt to the perturbation may be related to the fact that they rely more on their somatosensory feedback than on their auditory feedback during speech production (Lametti, Nasir, & Ostry, 2012). Of course, it would remain to be assessed by other means that the pattern of superposition of phonological units displayed in this study for subjects who adapt can be extended to all subjects, even those who do not adapt and hence cannot display learning transfer processes.

Whether observed transfer patterns can be extended to all speakers in a given language is related to more general questions about idiosyncrasies, plasticity, and evolution of phonological representations in the course of life. First, this study is limited to the selection of tested speakers and possibly to the limited range of their age or sociocultural provenance. Although there is no indication in previous work that educational background or socioeconomic status has an effect on speakers' magnitude of adaptation or transfer, as these phenomena are considered to be pure sensorimotor ones, further work could address this kind of question. We cannot also discard at this stage the possibility that the selection of appropriate units driving the definition of motor commands differs among speakers and that our global data display a mixture of different individual behaviors. However, our experimental design does not allow us to assess this hypothesis. More importantly, the theory of developmental phonology (Vihman & Croft, 2007), mainly based on the analysis of infants' babbling and their first produced words, suggests an evolution of speech units over the course of development (see also Tilsen, 2016). According to this theory, words would be the initial basic units of speech production in infants. Syllables and phonemes would emerge at a later stage from the observation of similarities in the production of the first words (Cristia & Hallé, 2012). Words, syllables, and phonemes would hence coexist as encoding units in adults. This suggests that auditory–motor learning transfer could provide an adequate paradigm to study the dynamics of the development of the phonological network along life. It also provides a natural tool for dealing with plasticity associated with clinical applications.

Clinical Outlook

Our results suggest that transfer of learning can occur at the level of several linguistic units (e.g., the word or

the utterance, the syllable, and the phoneme) in parallel. While this study concerns artificial learning driven by a perturbation, treatments of apraxia focus on (re)learning to speak (Maas et al., 2008). If transfer of learning works similarly in these two different situations (which needs to be validated), then the results of this study suggest that several linguistic units may be targeted in rehabilitation programs. A recent study of four patients with AOS focused on the transfer of learning at different levels in terms of linguistic units (Aichert & Ziegler, 2013). The authors observed some transfer from a training syllable to the same syllable embedded in disyllabic words, which is consistent with our results. On the contrary, they did not observe any transfer from the production of an isolated phoneme (a consonant) to the same consonant embedded in a syllable, whereas we found transfer at the level of the phoneme. This may point out the importance of the context in which the targeted unit is embedded during training. The influence of the context on transfer could be further investigated using the paradigm of auditory-motor learning. In addition, the position effect observed in the current study may draw the attention of speech therapists on the possible influence of the position of the targeted syllable in the word when selecting specific training utterances for their patients. Overall, this study introduces a method to investigate transfer of learning at various linguistic levels in typical speakers. This could prove useful to fine-tune some aspects of the treatments of patients with AOS.

Conclusion

The purpose of this study was to investigate the role of several speech units in phonological encoding at the interface between the linguistic input and speech articulation. The transfer of sensorimotor learning from a vowel in a specific context to the same vowel in a different context has been regarded as evidence for the existence of a common representation shared between the training and test utterances, namely, the phoneme (Houde & Jordan, 1998). The present experiment extended this result to other speech units (the syllable and the utterance) by contrasting the transfer magnitude at different levels: (a) to the same vowel in a different CV syllable, (b) to the same syllable in a different utterance, and (c) to the same utterance. The results suggest that words, syllables, and phonemes are all part of the interface with speech articulation.

In addition, the transfer was greater to the first than to the second syllable of the transfer utterance. This position effect provides new perspectives in regard to the serial control and sequential encoding of speech production.

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013, Grant 339152) awarded to Jean-Luc Schwartz.

References

- Aichert, I., & Ziegler, W. (2013). Segments and syllables in the treatment of apraxia of speech: An investigation of learning and transfer effects. *Aphasiology*, 27(10), 1180–1199. <https://doi.org/10.1080/02687038.2013.802285>
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, 17, 97–110. <https://doi.org/10.1371/journal.pone.0069107>
- Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, 8, Article 208. <https://doi.org/10.3389/fnhum.2014.00208>
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–290.
- Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of the 8th ISSP* (pp. 65–68). Rocquencourt, France: INRIA.
- Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America*, 128(4), 2033–2048. <https://doi.org/10.1121/1.3479539>
- Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205–235. <https://doi.org/10.1016/j.cognition.2005.01.009>
- Cristia, A., & Hallé, P. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 11–38). Frankfurt am Main, Germany: Lang.
- Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders*, 40(4), 467–491. <https://doi.org/10.1080/13682820500126049>
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321.
- Fant, G. (1960). *Acoustic theory of speech production* (Vol. 2). Paris, France: De Gruyter Mouton.
- Ferrand, L., & Segui, J. (1998). The syllable's role in speech production: Are syllables chunks, schemas, or both? *Psychonomic Bulletin & Review*, 5(2), 253–258. <https://doi.org/10.3758/BF03212947>
- Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 723(35), 708–723. <https://doi.org/10.1006/jmla.1996.0037>
- Frank, A. F. (2010). *Integrating linguistic, motor, and perceptual information in language production* (Doctoral dissertation). University of Rochester, NY.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Hawco, C. S., & Jones, J. A. (2010). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*, 1276, 131–139. <https://doi.org/10.1016/j.brainres.2009.04.033>.Control

- Hoffmann, L. A., & Sober, S. J. (2014). Vocal generalization depends on gesture identity and sequence. *Journal of Neuroscience*, 34(16), 5564–5574. <https://doi.org/10.1523/JNEUROSCI.5169-13.2014>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310. <https://doi.org/10.1092-4388/02/4502-0295>
- Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14(1), 147–172. <https://doi.org/10.1515/prbs.2002.002>
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27), 9351–9358. <https://doi.org/10.1523/JNEUROSCI.0404-12.2012>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232. [https://doi.org/10.1016/S1364-6613\(99\)01319-4](https://doi.org/10.1016/S1364-6613(99)01319-4)
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Liu, C., Cripe, T. P., & Kim, M. (2009). Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. *Molecular Therapy*, 18(9), 1724–1730. <https://doi.org/10.1038/mt.2010.127>
- Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, 17(3), 277–298.
- MacNeillage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465), 527–531. <https://doi.org/10.1126/science.288.5465.527>
- Manoiloff, L., Segui, J., & Hallé, P. (2016). Subliminal repetition primes help detection of phonemes in a picture: Evidence for a phonological level of the priming effects. *The Quarterly Journal of Experimental Psychology*, 69(1), 24–36. <https://doi.org/10.1080/17470218.2015.1018836>
- Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory–motor learning in speech production. *Neuroscience Letters*, 591, 25–29. <https://doi.org/10.1016/j.neulet.2015.02.012>. Feedback
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29(5), 524–545. [https://doi.org/10.1016/0749-596X\(90\)90050-A](https://doi.org/10.1016/0749-596X(90)90050-A)
- Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: Turning noise into signal pollution. *The American Naturalist*, 173(1), 119–123. <https://doi.org/10.1086/593303>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516–524. <https://doi.org/10.3758/BF03195598>
- Nooteboom, S., & Quené, H. (2015). Word onsets and speech errors. Explaining relative frequencies of segmental substitutions. *Journal of Memory and Language*, 78(C), 33–46. <https://doi.org/10.1016/j.jml.2014.10.001>
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>
- Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, 31(1), 4–18. <https://doi.org/10.1080/23273798.2015.1081703>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2007). *nlme: Linear and nonlinear mixed effects models*. Retrieved from <https://CRAN.R-project.org/package=nlme>
- Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297. <https://doi.org/10.1121/1.2173514>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from <http://www.R-project.org/>
- Recasens, D. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, 102(1), 544–561.
- Reilly, K. J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology*, 118(5), 2925–2934. <https://doi.org/10.1152/jn.00702.2016>
- Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *The Journal of Neuroscience*, 31(7), 2657–2662. <https://doi.org/10.1523/JNEUROSCI.6020-10.2011>
- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717. <https://doi.org/10.1152/jn.00773.2011>
- Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, 14(2), 173–200. <https://doi.org/10.1080/016909699386338>
- Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom & S. E. G. Öhman (Eds.), *Frontiers of speech communication research* (pp. 217–229). New York, NY: Academic. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=mzh&AN=1983016885&lang=fr&site=eds-live>
- Sevold, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34(6), 807–820. <https://doi.org/10.1006/jmla.1995.1035>
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42(1–3), 213–259. [https://doi.org/10.1016/0010-0277\(92\)90044-I](https://doi.org/10.1016/0010-0277(92)90044-I)
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 41–55. [https://doi.org/10.1016/S0022-5371\(79\)90554-1](https://doi.org/10.1016/S0022-5371(79)90554-1)
- Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55, 53–77.

- Tourville, J. A., Reilly, K. J., & Guenther, F. H.** (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39, 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Trautmüller, H.** (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, 69(5), 1465–1475.
- Välimaa-Blum, R.** (2009). The phoneme in cognitive phonology: Episodic memories of both meaningful and meaningless units? *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, 2. <https://doi.org/10.4000/cognitextes.211>
- Vihman, M., & Croft, W.** (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, 45(4), 683–725. <https://doi.org/10.1515/LING.2007.021>
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H.** (2007). Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. <https://doi.org/10.1121/1.2773966>
- Yates, A. J.** (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3), 213–232. <https://doi.org/10.1037/h0044155>

4. Apprentissage sensorimoteur, transfert, et développement des unités phonologiques

4.1 Introduction

Au travers de la revue de littérature de la section 5 du premier chapitre, nous avons vu que la lecture peut avoir une influence importante sur le traitement de la parole en perception et en production. Réciproquement, le langage oral, notamment à travers l'accroissement des vocabulaires productif et réceptif, permet de développer la conscience phonologique, une compétence étroitement liée à l'apprentissage de la lecture. Les représentations phonologiques, qui jouent un rôle fondamental dans l'acquisition et la maîtrise de ces trois compétences, sont probablement un des éléments clés à la base de leurs interactions.

Dans ce chapitre 4, nous nous intéressons aux interactions entre lecture, conscience phonologique et représentations phonologiques sous-tendant la parole, au travers de deux études expérimentales, présentées chacune sous forme d'article. La première étude a été initialement conçue dans la perspective de préparer la seconde. Elles abordent néanmoins la problématique sous deux angles très différents. Une représentation schématique des thèmes abordés, adaptée du chapitre 1, est proposée en Figure 4.1.

La première étude teste si le fait de présenter au sujet le mot à prononcer sous forme écrite ou sous la forme d'une image a une influence sur les phénomènes d'apprentissage auditorimoteur et son transfert. Il s'agit donc d'évaluer une potentielle influence de l'activation de représentations orthographiques. Ce travail a fait l'objet d'un acte de conférence internationale avec comité de lecture.

Caudrelier, T., Perrier, P., Schwartz, J.-L., & Rochet-Capellan, A. (2018). Picture naming or word reading: does the modality affect speech motor adaptation and its transfer? In *Proceedings of Interspeech 2018* (pp. 956–960).

La seconde étude explore l'évolution des représentations phonologiques, en reprenant le paradigme d'adaptation/transfert et en l'adaptant à des enfants entre deux tranches d'âge situées avant et après l'apprentissage de la lecture. Elle s'intéresse dans cette perspective aux liens entre ces représentations et la conscience phonologique. Ce travail fait l'objet d'un article court (« Brief communication ») publié par le journal *Cognition* :

Caudrelier, T., Ménard, L., Perrier, P., Schwartz, J., Gerber, S., Vidou, C., & Rochet-Capellan, A. (2019). Transfer of sensorimotor learning reveals phoneme representations in preliterate children. *Cognition*, 192(11), 103973. <https://doi.org/10.1016/j.cognition.2019.05.010>

Cette seconde étude a aussi fait l'objet d'une communication à la conférence internationale ICPC (Internationale Child Phonology Conference) à Chania, en 2018 :

Caudrelier, T., Ménard, L., Schwartz, J., Perrier, P., & Rochet-Capellan, A. (2018). Is transfer of learning related to phonological development? An acoustic study in 4 year-old and 8 year-old children. In *International Child Phonology Conference*.

L'article étant dans un format court, il ne présente pas tous les détails pouvant être utiles à la discussion générale de cette thèse. Des résultats complémentaires sont donc présentés dans ce chapitre, à la suite de l'article, en section 4.4.

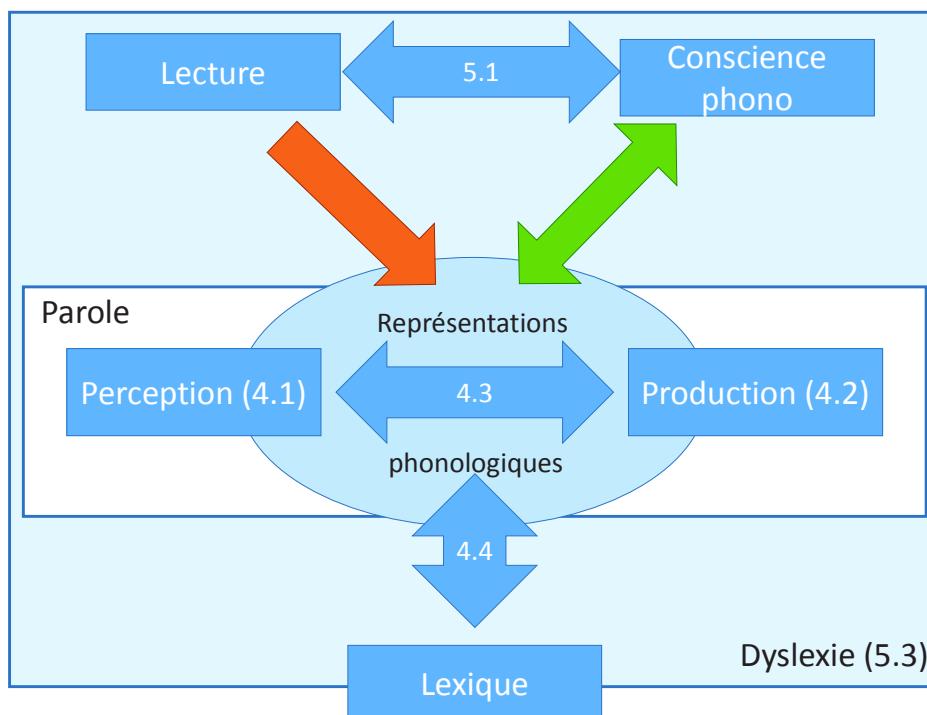


Figure 4-3: Schéma des relations entre parole, lecture et conscience phonologique, adapté du chapitre 1. La flèche orange représente la question traitée par l'étude 1. La flèche verte représente les relations investiguées dans l'étude 2.

4.2 Etude 2 - Dénomination d'images ou lecture de mots : la modalité affecte-t-elle l'adaptation auditorimotrice et son transfert ?

Pour tester l'adaptation auditorimotrice et son transfert chez des sujets ne maîtrisant pas la lecture, en particulier chez des enfants de moins de 6 ans, le stimulus utilisé ne peut pas être un mot écrit, comme chez l'adulte lettré, mais plutôt une image. Cependant, lire un mot ne met pas en jeu les mêmes traitements que nommer une image, comme cela a été montré avec un paradigme d'amorçage, ou *priming* (Ferrand, Grainger, & Segui, 1994). Ces différences pourraient reposer sur l'activation en temps réel de représentations orthographiques lors de la lecture d'un mot, mais pas lors de la dénomination d'images. Dans cette étude, nous nous sommes demandés si la modalité de présentation du stimulus à prononcer, une image ou un mot écrit, avait une influence sur l'apprentissage auditorimoteur et son transfert.

Nous avons estimé l'adaptation sur le stimulus d'entraînement /pe/ en fonction des deux modalités de présentation, et son transfert au mot /epe/ ('épée') dans la même modalité ou non, selon un design expérimental mixte de type 2*2. De même, le post-effet sur /pe/ a été évalué selon qu'il est présenté dans la même modalité que l'adaptation, ou non.

Les résultats diffèrent selon que l'on s'intéresse à l'adaptation et à sa rétention via le post-effet, ou bien à la généralisation de l'adaptation via le transfert. La modalité n'a pas d'influence sur l'adaptation ni le post-effet. En revanche, l'amplitude du transfert dépend de l'interaction entre la syllabe (1^{ère} ou 2^{ème} syllabe de /epe/) et la modalité du stimulus de transfert.

Ainsi, l'influence de la modalité sur l'amplitude du transfert pourrait être le signe d'une activation différente des représentations phonologiques, sous l'influence des représentations orthographiques, lors de la tâche. Plus généralement, cette observation indique que les représentations sensorimotrices des unités phonologiques mises en jeu sont sous l'influence de la modalité de présentation des stimuli. D'autre part, l'absence d'effet sur l'adaptation et le post-effet, et l'existence d'un transfert significatif avec la tâche de dénomination d'images, a permis de valider notre projet d'utiliser des images pour l'étude suivante portant sur le développement de l'enfant.



Picture naming or word reading: does the modality affect speech motor adaptation and its transfer?

Tiphaine Caudrelier¹, Pascal Perrier¹, Jean-Luc-Schwartz¹, Amélie Rochet-Capellan¹

¹Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France

* Institute of Engineering Univ. Grenoble Alpes

Tiphaine.caudrelier@gipsa-lab.fr, Amelie.rochet-capellan@gipsa-lab.fr

Abstract

Auditory-motor adaptation and transfer paradigms are increasingly used to explore speech motor control as well as phonological representations underlying speech production. Auditory-motor adaptation is generally assumed to occur at the sensory-motor level. However, few studies suggested that linguistic or contextual factors such as the modality of presentation of stimuli influences adaptation.

The present study investigates the influence of the modality of stimuli presentation (written word vs. a picture representing the same word) on auditory-motor adaptation and transfer. In this speech production experiment, speakers' auditory feedback was altered online, inducing adaptation. We contrasted the magnitude of adaptation in these two different modalities and we assessed transfer from /pe/ to the French word /epe/ in the same vs. different modality of presentation, using a mixed 2*2 subject design.

The magnitude of adaptation was not different between modalities. This observation contrasts with recent findings showing an effect of the modality (a written word vs. a go signal) on adaptation. Moreover, transfer did occur from one modality to the other, and transfer pattern depended on the modality of transfer stimuli. Overall, the results suggest that picture naming and word reading rely on sensory-motor representations that may be linked to contextual (or surface) characteristics.

Index Terms: auditory perturbations, adaptation, transfer, modality, picture naming, phonological units

1. Introduction

Adaptation to formants perturbation is increasingly used to investigate speech motor control in adults and children, including typical and atypical speakers [1]–[3]. This paradigm consists in applying a real-time systematic perturbation to one or several formants in speakers' auditory feedback. With training, the perturbation induces a change in speakers' pronunciation, called adaptation. This adaptation may or may not transfer to untrained speech material [4]. In most auditory-motor adaptation experiments, utterances were prompted by written words, but studies involving children or illiterate adults rather used pictures [5]. It is however unclear if the modality of presentation (written word vs. picture) affects auditory-motor adaptation and its transfer to other utterances. In other words, is the access to phonological representations and their associated sensory-motor characterizations mediated in the same way according to whether the stimuli are written-words or pictures?

Effects of non-linguistic properties of stimuli (or surface characteristics effects) have been found in research on memory

and language [6]. For example, the typography of stimuli (hand-written in uppercase vs. typed in lowercase) was shown to influence priming effect in a word-fragment completion task [7]. In speech production, Ferrand et al. [8] found that written word primes facilitated both word and picture naming. Moreover the authors showed that priming effects in picture naming relied mostly on phonological representations of whole words, while priming in word reading was primarily sensitive to orthographic overlap between the prime and the target [8]. Hence, reading vs. picture naming may activate different processes and representations [9].

The effect of modality observed in priming studies questions the use of pictures vs. written words as neutral factors in auditory-motor adaptation studies, and in particular studies using formant perturbations and analyzing transfer of adaptation. Indeed, the priming paradigm and the auditory-motor adaptation transfer paradigm have in common that they aim at revealing common speech representations between different stimuli, words or situations. These words are the "prime" and the "target" in the former, whereas the latter reveals common representations between a training word and a testing word. In auditory-motor adaptation studies, speakers produce one or several training word(s) while one or several formants are shifted in real time in their auditory feedback [4], [10], [11]. Over repetitions, most speakers adapt to the perturbation: they learn to partially compensate for it by changing their production in the opposite direction of the perturbation (e.g., in the production of "head", when F1 and F2 are shifted toward "had", production is adapted toward "hid"). The pronunciation change persists at least transiently when the perturbation is removed either for the pronunciation of the training material (after-effect) or for the pronunciation of other speech material (transfer effect). The study of transfer profiles was introduced as a way to question the nature of speech production representations [4], [11]–[15]. More specifically, contrasting transfer at the level of different linguistic units (utterances, syllables, and phonemes) was used to reveal speech representations interfacing with speech articulation [15].

Adaptation to formants perturbation is assumed to change auditory-motor representations. Yet several studies have shown an effect of linguistic factors on this adaptation. In particular, the magnitude of adaptation is influenced by the lexical status of the produced word and heard word, with larger adaptation for pseudo-words shifted to real-words than for real-words shifted to pseudo-words [16]. The frequency of the word which speakers are expected to pronounce to compensate for the perturbation also influences adaptation, with lower amplitude of adaptation if the articulatory target is more frequent [17]. Moreover, in a recent study, Alsius et al. [18] contrasted adaptation occurring when the prompt to produce the training word "head" was a linguistic stimulus (a written word or a

spoken word) vs. a go signal (a cross or a tune). Results suggested that linguistic prompts induced more adaptation than non-linguistic ones, regardless of the sensory modality, which was interpreted as evidence that linguistic prompts tend to define more precise speech goals than non-linguistic ones.

All together, these various studies suggest that representations involved in speech articulation might be related both to linguistic and surface factors. To further investigate the link between these different levels, the present study contrasts the magnitude of adaptation to formants perturbation on the French pseudo-word /pe/ in two different modalities of presentation: a written word (W) vs. a picture (P) of a fictional character. We then evaluated the interaction between linguistic and surface effects by testing transfer of adaptation to the French word /epe/ (meaning “sword”) and after-effects in the same modality than training (word-to-word (WW) and picture-to-picture (PP)) vs. in different modalities (word-to-picture (WP) and picture-to-word (PW)).

2. Method

2.1. Predictions

Our main hypothesis is that access to orthographic representations interferes with the specification of the phonological content of stimuli. Thus it would predict (1) different adaptation when speakers are trained with a written word (W) than with a picture (P) and (2) different transfer depending on the modality of presentation of the transfer stimulus. In addition, the role of surface characteristics would predict (3) more transfer and after-effect in the same modality as training (WW and PP conditions) than in the other modality (WP and PW conditions).

2.2. Participants, task and auditory feedback perturbation

Twenty-six native speakers of Metropolitan French ranging in age from 18 to 31 years old (mean=23, 11 women), were tested in a soundproof room. All reported no history of speech, language or hearing disorders. They were naïve to the purpose of the experiment.

Talkers were prompted to read words or name pictures displayed on a monitor, while wearing headphones with a built-in microphone (Sennheiser HME 26-II-600). They were instructed to speak aloud in a natural way, without shouting or whispering. They heard their own voice at about 70 dB mixed with speech shaped noise at 50 dB in the headphones.

During the training phase (see experimental procedure described below), the first two formants were shifted online using the Audapter system [19]. F1 was increased by 27% and F2 was decreased by 10%, making the vowel /e/ closer to /ɛ/ in

speakers’ auditory feedback. The perturbation triggered a non-disruptive 14ms delay.

2.3. Stimuli, experimental design and procedure

Stimuli were either written words or pictures (see Fig.1). As the training word “pé” (/pe/) is not a real French word, a fictional character was used to represent the pseudo-word /pe/ (participants were told that the character was named /pe/). /epe/ was elicited by a drawing of a sword (“épée” means sword in French).

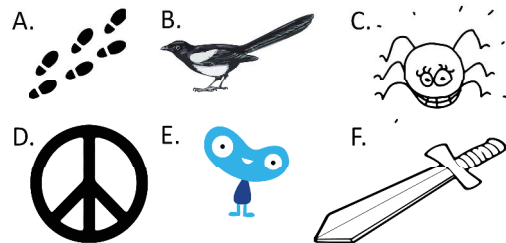
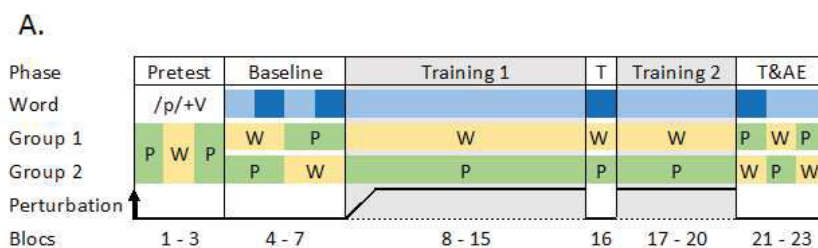


Figure 1: Pictures A, B, C, D, E, F prompted /pa/ (footprint), /pi/ (magpie), /pu/ (louse), /pe/ (peace), the training word /pe/ (a fictional character called /pe/ in the experiment), and the test word /epe/ (sword), respectively. A, B, C and D were used during the pretest only.

The experimental design and procedure are illustrated in Figure 2. The experiment consisted of 23 blocks of 20 trials, divided into 6 phases. Between each block, participants could rest and they had to press a key to continue the task. In the pretest, to explore their vocalic triangle, speakers were prompted to pronounce a set of syllables of the form /p/+Vowel, consisting in four real words /pa/, /pi/, /pu/, and /pɛ/ and one pseudo-word (/pe/). The /p/+V syllables were prompted alternately by corresponding written words or pictures (see Figure 1). In the baseline (blocks 4 to 7), speakers pronounced 40 times the training utterance (/pe/) and the testing utterance (/epe/) without any perturbation. Half of the utterances were prompted by pictures, and the other half by written words (presented in successive blocks, as specified in Figure 2). In the training phase (blocks 8 to 15), both groups pronounced the training utterance /pe/ while the perturbation was applied to their auditory feedback. The stimuli were the written word “pé” in group 1 and the picture associated with /pe/ in group 2. The first block after training (block 16) tested the transfer to /epe/ in the same modality (word in group 1 and picture in group 2, which corresponds to conditions WW and PP respectively). As the transfer was assessed with no perturbation, adaptation could progressively decay. Therefore, speakers underwent a second training phase (block 17 to 20) to adapt again to the perturbation. Transfer to /epe/ was then assessed in the other



B.

Phase	Transfer		After-effect	
	1	2	1	2
Group 1	WW	WP	WW	WP
Group 2	PP	PW	PP	PW

Figure 2: A. Experimental procedure schematic describing stimuli and perturbation by group over the course of the experiment. B. Experimental conditions by group and by experimental phase

modality (picture in group 1 and word in group 2; WP and PW respectively). Finally, after-effect on /pe/ was evaluated first in the same modality (word in group 1 and picture in group 2, WW and PP respectively) and then in the other modality (picture in group 1 and word in group 2, WP and PW respectively).

2.4. Data analyses

Formants F1 and F2 were evaluated in a 30ms window in the stable part of each vowel. The difference F2-F1 was computed in Hz to assess the magnitude of adaptation, transfer and after-effect as it accounts for changes in both formants. In each speaker, F2-F1 was expressed as a percentage of change compared to the baseline of the same syllable, associated with the exact same stimulus in the same modality of presentation.

Adaptation was evaluated in the last block of each training phase. For each subject, a t-test determined whether the subject significantly adapted to the perturbation or not.

To assess the magnitude of adaptation, transfer and after-effect, we built several Linear Mixed Models (LMM) using the lmer function from the lme4 package of R software. A backward deletion approach [20] enabled to select the best linear mixed model based on likelihood-ratio test. In all LMMs, the change in (F2-F1) was the variable to explain, and the random factor was the speaker. Fixed factors considered were (a) experimental phase and modality of training in the adaptation model; (b) modality of training, modality of transfer and syllable in the transfer model; (c) modality of training and block in the after-effect model. Post-hoc tests were adjusted for multiple comparisons, using the glht function from the multcomp package of R [21].

3. Results

As a result of the auditory feedback perturbation, 11 subjects out of 13 significantly adapted to the perturbation in each group by increasing the difference between their first two formants (F2-F1) at the end of the first training phase in comparison with the baseline. Non-adapted speakers were included in the analysis of adaptation in Section 3.1 (as in previous work on adaptation [18]) but then excluded from the analyses of transfer and after-effect in Sections 3.2 & 3.3, since these phenomena are by construction based on adaptation [15]. In addition, two subjects (one per group) were excluded from all results because they did not follow the instructions properly (one produced the wrong word in the baseline, and the other missed most trials because he was falling asleep).

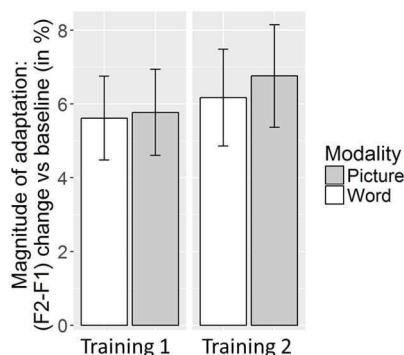


Figure 3: Magnitude of adaptation by modality of training (written words vs. pictures) and by training phase. Error bars represent standard errors.

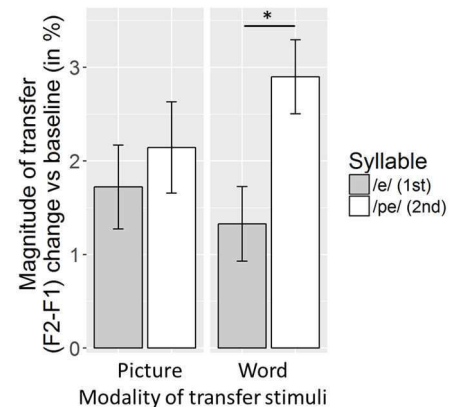


Figure 4: Magnitude of transfer by modality and by syllable of /epe/. Modalities of adaptation are combined. Error bars represent standard errors. * $p < 0.05$

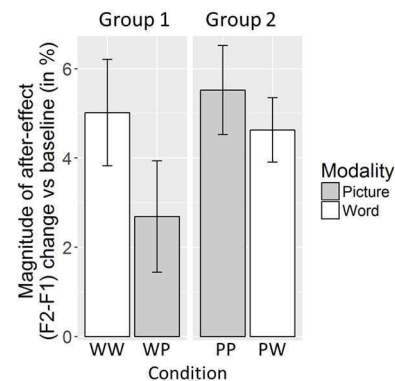


Figure 5: Magnitude of after-effect averaged by condition. Error bars represent standard errors.

3.1. Adaptation

Figure 3 displays the magnitude of adaptation in both groups and both training phases. The average adaptation was $6.1\% \pm 1.23$, with an increasing trend between the first and second training ($+0.77\%$). The best LMM fitting data did not depend on experimental phase although its effect was almost significant ($\text{Chisq}(1)=3.43$; $p=0.064$). The modality of training (word vs. picture) had no significant effect ($\text{Chisq}(1)=0.05$; $p=0.8$).

When excluding non-adapted speakers, the increase between training 1 and training 2 stayed not significant ($\text{Chisq}(1)=3.59$; $p=0.058$).

3.2. Transfer

We analyzed transfer magnitude for the word /epe/, by syllable, by modality of adaptation and by modality of transfer. Some transfer seemed to occur in all conditions. The best LMM fitting transfer data showed that modality of adaptation played no significant role while transfer depended on the interaction between syllable and modality of transfer (see Fig. 4). Post hoc comparisons revealed that transfer was significantly lower in

syllable /e/ than in /pe/ (i.e. first vs. second syllable) when /epe/ was prompted by the word “épée” (i.e. in conditions PW or WW; $z=-2.824$; $p=0.032$) but not when prompted by the picture (i.e. in conditions PP or WP: $z=-0.812$; $p=0.917$). Post hoc tests also confirmed that transfer magnitude was significantly above 0 in all conditions of adaptation and transfer ($p<0.001$).

3.3. After-effect

After the second transfer block, when perturbation is stopped, after-effect on the training utterance /pe/ is assessed on the same modality as training and then on the other modality (see Fig. 5). In both groups, after-effect seems to decay between first and second block of after-effect. The best LMM had block as the only factor (Chisq(1)=6.03; $p=0.014$). Hence globally, WW and PP have higher after-effects than WP and PW.

4. Discussion

In the present study, we assessed the effect of the modality of presentation of stimuli (written words vs. pictures) on sensory-motor adaptation to auditory perturbations. More specifically, we introduced the change in modality between adaptation and its transfer and after-effect, as a way to question the role of “surface effects” [6] in auditory-motor adaptation. As in our previous work [15], we argued that the study of transfer might highlight the nature of the link between phonological content and sensory-motor representations in speech production.

The first question concerned the possible interference of orthographic factors in the phonological specification when training stimuli were written words as opposed to pictures. Results show significant and perceivable [22] adaptation, but no significant difference in the adaptation between modalities. This observation does not corroborate recent findings suggesting that written or spoken words induce more auditory-motor adaptation than non-linguistic “go” signals (such as a visual cross or a tune) [18].

The absence of significant results could have several explanations. It could be that the number of participants in the present study was not large enough to enable significant results to be found (Alsius et al’s study [18] included 64 participants vs. 26 participants in the present study). However, there was absolutely no trend related to modalities in our adaptation data, so this interpretation seems unlikely to us. Alternatively, the absence of effect of modality on adaptation could be explained by the nature of the stimulus. Participants were told that the fictional character associated with the training utterance was called /pe/, whereas in [18] the visual go signal was a cross, that was not related to the word. If a name is explicitly given to the picture (or if the picture explicitly refers to a word), it may lead to a better specification of speech goals than in the case of a cross, and then lead to more adaptation. The influence of the nature of the link between the picture and its associated utterance could be investigated in further research.

Transfer was then assessed from /pe/ to the real word /epe/. We tested whether the orthographic partial overlap between the written words (“pé” and “épée”) could interfere with transfer (prediction 2) and investigated whether surface characteristics of stimuli were somehow linked to speech motor representations (in relation to prediction 3). The results show that interference between transfer and modality occurs only for transfer modality, but not for adaptation modality, showing that the effect of the syllable (/e/ vs. /pe/) depends on the modality of the test stimuli, regardless of the modality of the training

stimuli. As a matter of fact, the difference in transfer magnitude between /e/ and /pe/ in /epe/ is higher when the test stimuli are words (i.e. in WW and PW) than when they are pictures (in WP and PP), and this is not dependent on the modality used during adaptation. The transfer pattern observed with word stimuli is consistent with previous findings on disyllabic words [15]. Higher transfer was observed at a syllabic level (that is from /pe/ alone to /pe/ in /epe/) than at a phoneme level (that is from /pe/ alone to /e/ in /epe/). The interaction between test modalities and syllables may be interpreted in terms of surface characteristics effects [6]. At a phonological level, the second syllable is the same as the training utterance. The orthographic representation of this syllable may activate more intensely phonological representations, modified during training, than in the situation of picture naming. Furthermore, masked priming studies have shown that picture naming relied mostly on whole word phonological representations [8]. If pictures activated only holistic word representations [9], then transfer should not happen nor at a syllabic level neither at a phonemic one – which was not what we observed. The effect of the modality on the level at which transfer occurs the most (phoneme vs syllable vs word) could be investigated using a broader set of training and test words and/or pictures.

Finally, the possibility of a link between surface characteristics of stimuli and speech motor representations was tested in the after-effect. We actually found that there was significantly more after-effect in WW and PP than in WP and PW conditions. The results hence tend to suggest that after-effects are stronger when adaptation and test are made with the same modality. However, this interpretation should be nuanced, since this effect can be a simple effect of the block. Indeed, the after-effect gradually decays over repetitions since there is no more auditory perturbation when the after-effect is tested, and WP and PW were tested after WW and PP. This is a limitation of our experimental design, which will require further control by switching the order of same vs. other modalities in the protocol (that is WP then WW in group 1 and PW then PP in group 2). Interestingly, the after-effect seemed slightly lower in the WP condition than in the other conditions. Although this effect is not significant with the statistical power of the present study, it may pave the way to more extended work on this topic.

More generally, it is important to stress that even though the present results are not yet completely understood and in some cases inconclusive, they indicate that surface characteristics do matter in adaptation and transfer paradigms. Considering the increasing importance of this paradigm in recent research, the latter observation should be taken into account, especially in the design of experiments addressing adaptation or transfer in children or specific populations (such as illiterate adults).

5. Conclusions

Overall, the present observations display several effects that may be related to surface characteristics of stimuli. To the best of our knowledge, this study is the first to investigate modality effects in the transfer of auditory-motor adaptation. The results suggest that this topic could be insightful in the understanding of representations tied to speech articulation, and in particular to understand the level of abstraction of these representations.

6. Acknowledgements

The research leading to these results has received funding from the European Research Council under the European

7. References

- [1] L. Max and D. G. Maffett, "Feedback delays eliminate auditory-motor learning in speech production," *Neurosci. Lett.*, vol. 591, pp. 25–29, 2015.
- [2] H. Terband, F. Van Brenk, and A. Van Doornik-Van Der Zee, "Auditory feedback perturbation in children with developmental speech sound disorders," *J. Commun. Disord.*, vol. 51, pp. 64–77, 2014.
- [3] A. Daliri, E. A. Wieland, S. Cai, F. H. Guenther, and S. E. Chang, "Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter," *Dev. Sci.*, no. January 2016, pp. 1–11, 2017.
- [4] J. F. Houde and M. I. Jordan, "Sensorimotor adaptation in speech production.," *Science*, vol. 279, no. 1998, pp. 1213–1216, 1998.
- [5] E. N. MacDonald, E. K. Johnson, J. Forsythe, P. Plante, and K. G. Munhall, "Children's development of self-regulation in speech production," *Curr. Biol.*, vol. 22, pp. 113–117, 2012.
- [6] S. D. Goldinger, "Echoes of echoes? An episodic theory of lexical access.," *Psychol. Rev.*, vol. 105, no. 2, pp. 251–279, 1998.
- [7] H. L. Roediger and T. A. Blaxton, "Effects of varying modality, surface features, and retention interval on priming in word-fragment completion," *Mem. Cognit.*, vol. 15, no. 5, pp. 379–388, 1987.
- [8] L. Ferrand, J. Grainger, and J. Segui, "A study of masked form priming in picture and word naming," *Mem. Cognit.*, vol. 22, no. 4, pp. 431–441, 1994.
- [9] N. O. Schiller, "The effect of masked syllable primes on word and picture naming," *J. Acoust. Soc. Am.*, vol. 102, no. 5, p. 3136, 1997.
- [10] D. W. Purcell and K. G. Munhall, "Compensation following real-time manipulation of formants in isolated vowels.," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2288–2297, 2006.
- [11] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.*, vol. 122, no. 4, pp. 2306–2319, 2007.
- [12] S. Cai, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization.," *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 2033–2048, 2010.
- [13] A. Rochet-Capellan, L. Richer, and D. J. Ostry, "Nonhomogeneous transfer reveals specificity in speech motor learning," *J. Neurophysiol.*, vol. 107, no. 6, pp. 1711–1717, 2012.
- [14] T. Caudrelier, P. Perrier, J. L. Schwartz, and A. Rochet-Capellan, "Does auditory-motor learning of speech transfer from the CV syllable to the CVCV word?," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016, vol. 08–12–Sept, no. Cv, pp. 2095–2099.
- [15] T. Caudrelier, J.-L. Schwartz, P. Perrier, S. Gerber, and A. Rochet-Capellan, "Transfer of learning: what does it tell us about speech production units?," *J. Speech, Lang. Hear. Res.*
- [16] N. J. Bourguignon, S. R. Baum, and D. M. Shiller, "Lexical-perceptual integration influences sensorimotor adaptation in speech.," *Front. Hum. Neurosci.*, vol. 8, p. Article 208, 2014.
- [17] A. F. Frank, "Integrating Linguistic, Motor, and Perceptual Information in Language Production," University of Rochester., 2011.
- [18] A. Alsius, T. Mitsuya, N. Latif, and K. G. Munhall, "Linguistic initiation signals increase auditory feedback error correction," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 838–845, 2017.
- [19] S. Cai, M. Boucek, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/," in *Proceedings of the 8th ISSP*, 2008, pp. 65–68.
- [20] R. Mundry and C. L. Nunn, "Stepwise model fitting and statistical inference: turning noise into signal pollution.," *Am. Nat.*, vol. 173, no. 1, pp. 119–123, 2009.
- [21] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biometrical J.*, vol. 50, no. 3, pp. 346–363, 2008.
- [22] J. W. Hawks, "Difference limens for formant patterns of vowel sounds.," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1074–1084, 1994.

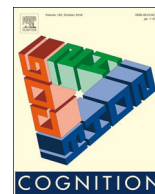
4.3 Etude 3 - L'apprentissage sensorimoteur révèle des représentations phonémiques chez les enfants pré-lecteurs

L'apprentissage de la lecture est étroitement lié au développement de la conscience phonologique, cette capacité de manipulation explicite des unités reposant sur des représentations phonologiques qualifiées d'*implicites*, et que nous considérons comme sensorimotrices. Dans cette étude, nous avons testé si les représentations sensorimotrices du phonème émergent avant l'acquisition de la lecture ou non.

Nous avons recruté trois groupes de participants : des enfants de 4-5 ans pas encore scolarisés, des enfants de 7-8 ans scolarisés sachant déjà lire, ainsi qu'un groupe d'adulte, utilisé comme contrôle. La conscience phonologique des enfants a été évaluée. Une perturbation en temps réel du retour auditif des participants a été appliquée lors de la production du mot d'entraînement /pe/, afin de provoquer un apprentissage auditorimoteur. Nous avons évalué le transfert d'apprentissage vers les deux /e/ de /epe/, c'est-à-dire à un niveau phonémique (le 1^{er} /e/) et syllabique (le 2^{ème} /e/, qui est inclut dans la syllabe /pe/). Le transfert au niveau phonémique a été utilisé comme un moyen de sonder l'existence de représentations phonémiques chez les participants.

Nous avons observé un transfert significatif chez les trois groupes d'âge, enfants préscolaires, scolaires et adultes. Nous avons également trouvé que les enfants s'étant significativement adaptés à la perturbation auditive avaient un meilleur score de conscience phonologique que les autres, et ce quel que soit le groupe d'âge. Ces résultats suggèrent que les représentations du phonème émergent avant l'acquisition de la lecture, et que ces représentations sensorimotrices pourraient être à la base du développement de la conscience phonologique.

De plus, un net contraste a été observé chez les enfants de 4-5 ans, entre l'amplitude du transfert sur /epe/ et celle du post-effet sur /pe/, contrairement aux autres groupes (voir section résultats complémentaires de ce chapitre). Cela suggère que les enfants pré-lecteurs se baseraient d'avantage sur des représentations holistiques du mot dans l'articulation de la parole, que ne le feraient les enfants en âge de lire, ou les adultes.



Brief article

Transfer of sensorimotor learning reveals phoneme representations in preliterate children

Tiphaine Caudrelier^a, Lucie Ménard^b, Pascal Perrier^a, Jean-Luc Schwartz^a, Silvain Gerber^a, Camille Vidou^b, Amélie Rochet-Capellan^{a,*}

^a Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, F-38000 Grenoble, France

^b Laboratoire de Phonétique, Université du Québec à Montréal, Center For Research on Brain, Language, and Music, Montreal, Quebec, Canada



ARTICLE INFO

Keywords:

Phoneme
Phonological awareness
Phonological development
Sensorimotor learning
Auditory perturbation
Transfer of learning

ABSTRACT

Reading acquisition is strongly intertwined with phoneme awareness that relies on implicit phoneme representations. We asked whether phoneme representations emerge before literacy. We recruited two groups of children, 4 to 5-year-old preschoolers (N = 29) and 7 to 8-year-old schoolchildren (N = 24), whose phonological awareness was evaluated, and one adult control group (N = 17). We altered speakers' auditory feedback in real time to elicit persisting pronunciation changes, referred to as auditory-motor adaptation or learning. Assessing the transfer of learning at phoneme level enabled us to investigate the developmental time-course of phoneme representations. Significant transfer at phoneme level occurred in preschoolers, as well as schoolchildren and adults. In addition, we found a relationship between auditory-motor adaptation and phonological awareness in both groups of children. Overall, these results suggest that phoneme representations emerge before literacy acquisition, and that these sensorimotor representations may set the ground for phonological awareness.

1. Introduction

During reading acquisition, children learn correspondences between graphemes (i.e. letters or sets of letters) and speech sounds, namely phonemes (Share, 1995). Literacy has been suggested to reorganize phonological representations (Pattamadilok, Knierim, Kawabata Duncan, & Devlin, 2010) in a way that may depend on the writing system (Wang, Wong, Wang, & Chen, 2017). Learning to read is closely related to phoneme awareness which is the capacity to explicitly identify and manipulate phonemes (Ziegler & Goswami, 2005). While phoneme awareness strongly predicts reading acquisition success (Melby-Lervåg, Lyster, & Hulme, 2012), learning to read may in turn be essential in the development of phoneme awareness (Morais, Cary, Alegria, & Bertelson, 1979). Thus there is probably a mutual facilitation between these two skills (Castles & Coltheart, 2004). Phoneme awareness is often claimed to rely on implicit phoneme representations (Fowler, 1991; Swan & Goswami, 1997). Therefore, a key question is: do these implicit phoneme representations emerge during reading acquisition in children, or prior to it?

In an emergentist view, the representations used by children when producing early words are usually word templates (Vihman & Croft, 2007). Then representations gradually become more fine-grained and

segmental (Cristia & Hallé, 2012). This evolution, interestingly related to vocabulary growth, may continue until middle childhood (Fowler, 1991; Nittrouer, Studdert-kennedy, & McGowan, 1989; Walley, Metsala, & Garlock, 2003). Although the term phonological representations – i.e. representations of phonological units, encompassing phonemes – has been much used, it is difficult to find measures probing phonological representations independently of other cognitive skills (Ainsworth, Welbourne, & Hesketh, 2016; Boada & Pennington, 2006).

We argue that phonological representations are essentially sensorimotor representations (Schwartz, Basirat, Ménard, & Sato, 2012; Tourville & Guenther, 2011). These representations involve both speech perception and production, associating sensory cues with motor characteristics, within categories. This view is consistent with the general developmental framework based on the dynamic systems theory (Smith & Thelen, 2003), proposing that representations are the manifestation of interactions between sensorimotor processes. This sensorimotor view of phonological representations may find support in the literature on dyslexia, which is often associated with phonological awareness deficits (Goswami, 2006). Indeed, in addition to various deficits in auditory processing (Boets et al., 2011; Noordenbos & Serniclaes, 2015) and implicit motor learning (Stoodley, Harrison, & Stein, 2006), dyslexia has been related to sensorimotor integration difficulties (Viana, Razuk,

* Corresponding author at: Speech & Cognition Department, GIPSA-lab, 11, rue des Mathématiques, GIPSA-lab, F-38000 Grenoble, France.
E-mail address: amelie.rochet-capellan@gipsa-lab.grenoble-inp.fr (A. Rochet-Capellan).

de Freitas, & Barela, 2013). Consistent with this finding, dyslexia has been associated to different sensorimotor responses to auditory perturbations compared to controls (Van Den Bunt et al., 2017; Van Den Bunt, Groen, van der Kleij, et al., 2018). Moreover, in typically developing preliterate children, reading-related skills (including phonological awareness) were found to be correlated with the properties of sensorimotor representations, as probed by an auditory-motor learning paradigm (Van Den Bunt, Groen, Frost, et al., 2018).

The auditory-motor learning paradigm consists of altering in real time the auditory feedback of a speaker producing an utterance (Houde & Jordan, 1998). For instance, the perturbation makes a vowel sound like another vowel. Speakers usually adapt to the perturbation in that they compensate for it by changing their pronunciation. This change persists even after feedback is returned to normal, revealing a change in the sensorimotor representation of the utterance, hereafter referred to as “the training word”. Whether this representation is defined at phonemes level can be probed by assessing generalization, also called transfer of learning (Cai, Ghosh, Guenther, & Perkell, 2010; Caudrelier, Schwartz, Perrier, Gerber, & Rochet-Capellan, 2018; Houde & Jordan, 1998; Rochet-Capellan, Richer, & Ostry, 2012). If the change in pronunciation transfers from the training word to another word produced under normal or masked feedback, it reveals the existence of a common representation shared by these words. Testing generalization as a means to probe representations has been used in the speech perception literature (McQueen, Norris, & Cutler, 2006; Mitterer, Reinisch, & McQueen, 2018; Mitterer, Scharenborg, & McQueen, 2013). To our knowledge, transfer of auditory-motor learning has never been studied in children in speech production.

In the present study, we investigated whether children relied on phoneme representations in speech production before the acquisition of reading, and before the emergence of phoneme awareness. We assessed transfer of adaptation at the phoneme level in preschoolers, in contrast with schoolchildren and adults. If such transfer occurs in preschoolers, then it can be assumed that phoneme representations in speech production emerge with speech experience rather than as a by-product of learning to read. We also tested whether robust auditory-motor adaptation is associated with higher phonological awareness in children, before and after reading acquisition, as shown by Van Den Bunt, Groen, Frost, et al. (2018) though only for preliterate children. Confirming such a link would reinforce the view that auditory-motor learning taps into phonological representations that underlie phonological awareness (PA).

2. Methods

2.1. Participants

Native speakers of Canadian French (N = 70) were included in this study, of which 29 preschoolers (4- to 5-year-old, mean = 4.9y,

sd = 0.6), 24 schoolchildren (7- to 8-year-old, mean = 8.0y, sd = 0.5), and 17 adults (mean = 25y, sd = 4). In addition to them, two preschoolers and one 8-year-old child participated but were excluded as they gave up or did not follow the task; another preschooler was excluded because of his high score in phoneme awareness (see in Section 3.3 for more information). Written informed consent was given by adult participants and parents of children, in agreement with Ethics statements (ethics approval number 390_e_2017, Institutional Review Board of the Université du Québec à Montréal).

Subjects had no reported speech, language or hearing disorders, and were naive to the purpose of the experiment. All participants had to perform the same production task, which was followed for all children by a PA test. The total duration did not exceed 20 min in preschoolers and 30 min in schoolchildren.

2.2. Phonological awareness test

During the PA test, all children performed a rhyme judgment task, a syllable deletion task in CVCV pseudo-words, and a phoneme deletion task in CVC. The second-grade group also performed phoneme deletions in CCV, syllable inversions in CVCV, and phoneme inversions in CV and VC. The rhyme judgment task was taken from a French Canadian test for preschoolers (Lefebvre & Sutton, 2008). The other tasks were taken from BELEC (Mousty, Leybart, Alegria, Content, & Morais, 1994), using blocks to represent phonological units in order to ease children’s comprehension of the task, as specified in Lefebvre and Sutton (2008).

2.3. Auditory-motor adaptation task

2.3.1. Equipment, task, and auditory perturbation

Speakers were instructed to name pictures presented on a monitor, speaking in a natural way. They heard their own voice mixed with speech shaped noise in headphones. During the training phase, the vowel /e/ was shifted towards /ε/ in the auditory feedback, by increasing F1 by 27% and decreasing F2 by 10% using Audapter (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Caudrelier, Schwartz, et al., 2018). Audapter formants prior values (fn1 and fn2 parameters) were adjusted for each speaker based on formants measured in prerecorded productions of /pe/.

2.3.2. Experimental procedure

The training utterance was the pseudoword /pe/ and the test utterance to assess transfer was the French word ‘épée’ (/epe/), which means “sword”. These choices were consistent with our previous study (Caudrelier, Schwartz, et al., 2018). Importantly, the test utterance was a real word well known by 4-year-old children and easy to prompt by a picture, to make sure that all children would remember it when they would have to produce it. Speakers were trained to provide the expected utterance from adequate pictures (a fictional character called

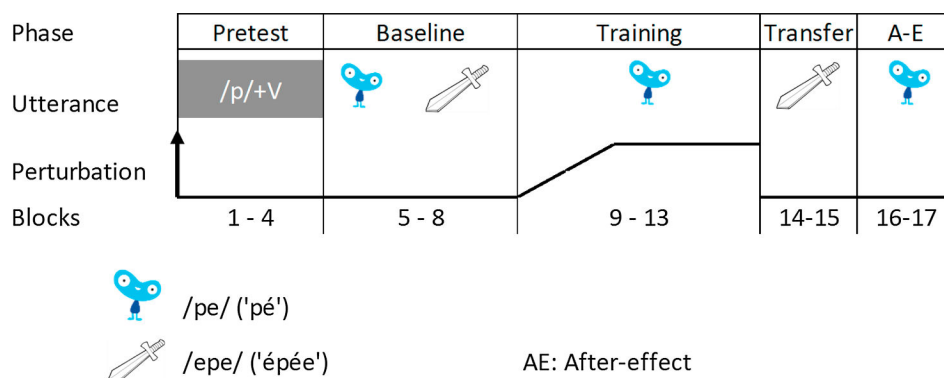


Fig. 1. Auditory-motor adaptation procedure.

“pé” /pe/ for training, and a picture of a sword for testing).

The procedure consisted of 5 phases (Fig. 1) made of blocks of 10 repetitions, as follows. (1) In the pretest (blocks 1 to 4, without perturbation) speakers pronounced monosyllabic words (/pa/, /pi/, /pu/, /pe/) and the training utterance (/pe/). (2) In the baseline (blocks 5 to 8, without perturbation) speakers produced /pe/ and /epe/ twenty times each. (3) In the training phase, the perturbation was gradually introduced during blocks 9 and 10, in four steps, and then held at its maximum (blocks 11 to 13), while speakers pronounced the training utterance /pe/. When the perturbation stopped, (4) speakers pronounced /epe/ in the transfer phase (blocks 14 and 15), and speakers repeated /pe/ again (5) in the after-effect phase (blocks 16 and 17).

2.4. Predictions

Consistent with previous observations (Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2018; Caudrelier, Schwartz, et al., 2018), we expected significant transfer at both syllabic and phonemic levels (i.e. in both /e/ of /epe/) in adults. We also predicted that: (H1) if phoneme representations emerge as a by-product of reading acquisition, then transfer pattern should be similar in 8-year-old children and in adults, whereas in 4-year-old children transfer should occur only at a syllabic level (from /pe/ to /pe/ in /epe/) but not at the phoneme level (to the first /e/ of /epe/). Conversely, (H2) if phoneme representations have already emerged in 4-year-olds (before reading acquisition) some transfer from /pe/ to /e/ should occur in preschoolers. Finally, (H3) if auditory-motor adaptation and transfer probe representations that set the ground for phonological awareness development, then we should find a relationship between adaptation and PA scores.

2.5. Measurements

Formants were measured in a 30-ms window in the middle of the vowel as in (Caudrelier, Schwartz, et al., 2018) and expressed in Barks (Schroeder, Atal, & Hall, 1979). Then, adaptation and transfer magnitudes were expressed as the percentage of change in (F2-F1) vs. the average (F2-F1) in the baseline, in the same utterance and the same speaker:

$$\left(\frac{(F2 - F1)_{\text{trial}}}{\text{mean}(F2 - F1)_{\text{baseline}}} - 1 \right) * 100$$

2.6. Statistical analysis

The analyses were carried out using R software (version 3.2.0 - Development Core Team, 2016).

2.6.1. Linear mixed models

Adaptation, transfer and after-effect magnitude were modeled by Linear Mixed Models (LMM) using a backward deletion approach. Potential autocorrelation and differences in variance across groups were taken into account (Liu, Cripe, & Kim, 2009). Post hoc tests were adjusted for multiple comparisons (Hothorn, Bretz, & Westfall, 2008). A first LMM modeled adaptation (block 13) and after-effect (block 16) with *age_group*, *phase* (training vs after-effect) and *trial* as fixed factors. Transfer of adaptation was assessed for adapted subjects, considering that if there was no adaptation there could not be transfer. Adapted speakers were detected by a non-parametric Wilcoxon test checking whether (F2-F1) in the last 20 trials of the training phase was significantly different from (F2-F1) in the 20 trials of the baseline. In another LMM which included only adapted speakers, *age_group*, *syllable* (1st vs 2nd syllable of /epe/) and *trials* were the fixed factors to model transfer of adaptation (blocks 14 and 15).

2.6.2. Relationship with phonological awareness

To assess whether sensorimotor behaviors are possible predictors of PA, the relationships between PA and (1) adaptation and (2) transfer were assessed with a Beta regression using *Betareg* function from R package *betareg* (Cribari-Neto & Zeileis, 2010). The regression modeled PA¹ with *age_group*, adaptation (adapters vs. non-adapters) and their interactions. Two variants of PA scores were evaluated: the score on the common part of the test between 4- and 8-year-olds to examine the effect of *age_group* without any bias related to the test, and the score on the whole test, which differed between age groups, to avoid ceiling effects in the 8-year-old group.

To assess the relationship between PA and transfer, PA score was modeled using the variables *age_group*, *syllable*, *transfer magnitude* and their interactions.

3. Results

3.1. Adaptation and after-effect

The average magnitude of adaptation and after-effect is shown in Fig. 2(a). It is similar across age groups. Indeed, the best model comprises only an interaction between phase and trial (Chisq(1) = 5.14; $p = 0.02$) but no *age_group* effect. Adaptation and after-effect magnitudes are significant, as shown by intercepts ($2.2\% \pm 0.5$; $z = 4.3$; $p < 0.001$ and $3.2\% \pm 0.5$; $z = 6.2$; $p < 0.001$). Adaptation is stable ($-0.02\%/trial \pm 0.02$; $z = -0.81$; $p = 0.9$) while after-effect significantly decreases over trials ($-0.09\%/trial \pm 0.03$; $z = -3.47$; $p = 0.003$).²

11 preschoolers, 10 schoolchildren and 12 adults significantly adapted to the perturbation (hereafter, the “adapted speakers”).

3.2. Transfer of learning

The dynamics of transfer magnitude is shown in Fig. 2(b). Transfer from /pe/ to /epe/ was equivalent across groups. The only fixed factor of the obtained model was trial (Chisq(1) = 14.2, $p < 0.0001$). Transfer was significantly above 0 at the beginning of the phase ($2.4\% \pm 0.5$; $z = 5.1$; $p < 0.001$). The slope was significantly negative ($-0.15\% \pm 0.03$; $z = -4.4$; $p < 0.001$). When analyzing preliterate children only, we obtained the same model and transfer was still significant at the beginning of the phase ($2.3\% \pm 0.8$; $z = 2.7$; $p = 0.01$). Hence transfer occurred in all age groups at both syllable and phoneme levels.

3.3. Phonological awareness

Table 1 shows PA scores by subtask in both groups of children. All preschoolers got phoneme deletion scores, related to phoneme awareness, close to 0 – actually all at 0 except 1 at 20% and 1 at 30%. The child excluded from the analyses (see Section 2.1) had a score at 70%.

3.3.1. Adapted vs non-adapted subjects

Fig. 3 shows PA scores in children who adapted to the perturbation vs. non-adapted children (preschoolers: $N = 29$; schoolchildren: $N = 24$). Regardless of the score taken into account (part 1 or total score), there was a significant effect of both *age_group* and *adaptation* (adapted vs. not adapted) factors, and no interaction between them. Multiple comparisons (with part 1 scores) show that: scores in 8-year-

¹ PA score was expressed in proportion of correct response, therefore ranging from 0 to 1. A transformation was applied to avoid 0 and 1 values as required by the Beta regression. The transformation is $(y \cdot (n - 1) + 0.5)/n$, with n being the number of participants and y the initial value (Smithson & Verkuilen, 2006).

² All full models output, as well as supplemental analyses including age as a continuous factor, can be found in the supplement.

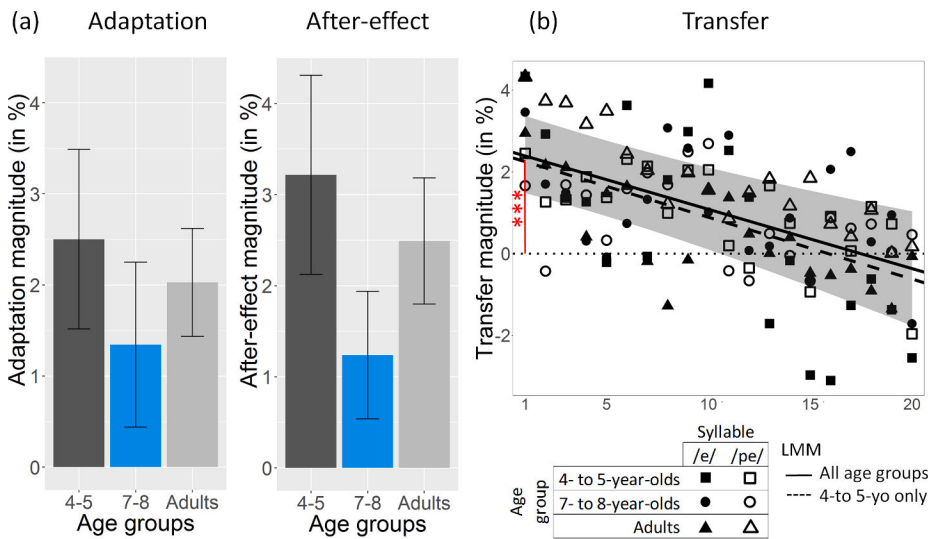


Fig. 2. (a) Average adaptation and after-effect magnitudes by age groups in all speakers (preschoolers: N = 29; schoolchildren: N = 24; adults: N = 17). Error bars represent standard errors. (b) Transfer magnitude by trial, by age group and by syllable in adapted speakers (preschoolers: N = 11; schoolchildren: N = 10; adults: N = 12). Shaded area shows the 95% confidence interval of the linear mixed model for all age groups, and *** ($p < 0.001$) is related to the intercept in the same model.

Table 1
Phonological awareness scores in % of correct responses by age group and by subtask (preschoolers: N = 29; schoolchildren: N = 24).

Phonological Awareness scores		Preschoolers		Schoolchildren	
(in % of correct responses)		mean	sd	mean	sd
Part 1	Rhyme judgement	60	25	94	7
	Syllable deletion	27	31	94	10
	Phoneme deletion in CVC	5	9	80	20
	Part 1 subtotal	33	19	90	9
Part 2	Phoneme deletion in CCV	/	/	78	21
	Syllable inversion	/	/	75	21
	Phoneme inversion	/	/	86	9
Total score		33	19	84	10

3.3.2. Transfer and PA

No relationship was found between transfer magnitude and PA scores in adapted speakers.

4. Discussion

Phoneme awareness, strongly intertwined with reading acquisition, relies on phoneme representations, which we assume to be sensorimotor representations (Schwartz et al., 2012). The present study was designed to determine whether these phoneme representations emerge before reading acquisition. These representations were probed using the paradigm of auditory-motor adaptation and transfer (Caudrelier, Schwartz, et al., 2018; Houde & Jordan, 1998).

Transfer at the phoneme level occurred in all age-groups, including the 4- to 5-year-old preschoolers, suggesting the existence of phoneme representations at this age. No evidence for age differences was found in the transfer pattern. In addition, as expected, phonological awareness (PA) tests indicate that preschoolers had some syllable awareness, but no or very little phoneme awareness (except one child that was excluded from the analysis). These observations suggest that phoneme sensorimotor representations emerge before the development of phoneme awareness, and before reading instruction starts, as already supported by Ventura, Kolinsky, Fernandes, Querido, and Morais (2007).

We found no significant difference in the average adaptation magnitude between groups. This suggests that the representations underlying auditory-motor adaptation are established early in development whereas phonological awareness continues to develop over childhood. There might though be an age effect that was not found for statistical power reason. This adaptation pattern differs from Van Den Bunt, Groen, Frost, et al. (2018) study reporting lower adaptation in 5-year-old than in 7-year-old children. However, our observations are consistent with two previous studies in 4-year-olds and older children which report equivalent adaptation magnitude across ages (MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012; Terband & van Brenk, 2015).

In both groups of children, average PA scores in adapted participants were significantly higher than average scores in non-adapted ones. This suggests that whether preschoolers adapt to the perturbation or not could be a predictor of later PA development, although a longitudinal study would be required to confirm this predictive role. These results extend the results by Van Den Bunt, Groen, Frost, et al. (2018) who found a relationship between adaptation magnitude and PA scores in preliterate children only.

The relationship observed between auditory-motor adaptation and PA at two different development stages may lead to several

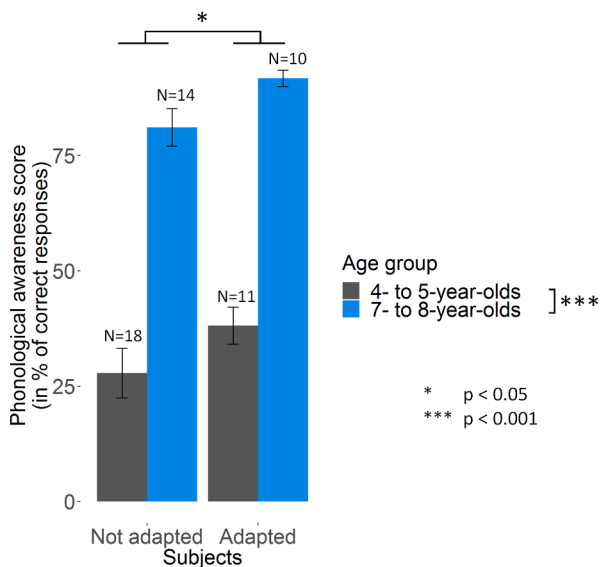


Fig. 3. Bar chart of phonological awareness total score by age group and whether children adapted to the auditory perturbation or not.

olds were significantly higher than in 4-year-olds ($z = 10.07$; $p < 0.0001$); and scores were higher in adapted speakers than in non-adapted speakers ($z = 2.25$; $p = 0.047$). Score difference between adapters and non-adapters was even larger when using total scores ($z = 2.68$; $p = 0.0147$).

interpretations. First, it may reveal a direct connection between PA skills and underlying sensorimotor representations which the paradigm taps into. This link may also be driven by a third factor that would influence both auditory-motor adaptation and PA. This factor could be categorical perception (CP), which has been shown to influence both reading abilities and auditory-motor learning (Niziolek & Guenther, 2013, see also Shiller & Rochon, 2014). Yet PA is associated with spoken vocabulary growth (Walley et al., 2003), suggesting that speech production – and not only perception – is involved in representations underlying PA development (see also Vihman, 2017). Since CP arises generally in the first year of life while adaptation mechanisms may establish between 2 and 4 years of age (MacDonald et al., 2012), we rather suggest that CP pave the way for the development of phoneme sensorimotor representations, which then serve as foundations for phoneme awareness acquisition.

The present study used an auditory-motor adaptation and transfer paradigm to probe phonological representations in children. The results provide evidence for the emergence of phoneme representations before the apparition of phoneme awareness and reading instruction. They also underline connections between sensorimotor representations and higher-level skills (Smith & Thelen, 2003). Training sensorimotor processing may be considered to support reading acquisition. Future studies could probe phonological representations in dyslexic people using the paradigm of auditory-motor learning transfer. This paradigm could also enable to explore how phonological representations of different languages relate to each other in multilingual speakers. Transfer of sensorimotor learning may thus turn out to be a valuable tool to study phonological representations and their development.

Declaration of Competing Interest

None.

Acknowledgement

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152), and from the Natural Sciences and Engineering Research Council of Canada (Grant #S-706186).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.5281/zenodo.2645764>.

References

- Ainsworth, S., Welbourne, S., & Hesketh, A. (2016). Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation. *Applied Psycholinguistics*, 37(4), 997–1023. <https://doi.org/10.1017/S0142716415000338>.
- Boada, R., & Pennington, B. F. (2006). Deficient implicit phonological representations in children with dyslexia. *Journal of Experimental Child Psychology*, 95(3), 153–193. <https://doi.org/10.1016/j.jecp.2006.04.003>.
- Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, 32(2), 560–570. <https://doi.org/10.1016/j.ridd.2010.12.020>.
- Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. *Proceedings of the 8th ISSP* (pp. 65–68).
- Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America*, 128(4), 2033–2048. <https://doi.org/10.1121/1.3479539>.
- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91(1), 77–111. [https://doi.org/10.1016/S0010-0277\(03\)00164-1](https://doi.org/10.1016/S0010-0277(03)00164-1).
- Caudrelier, T., Perrier, P., Schwartz, J.-L., & Rochet-Capellan, A. (2018). Picture naming or word reading: Does the modality affect speech motor adaptation and its transfer? In *Proceedings of Interspeech 2018* (pp. 956–960).

- Caudrelier, T., Schwartz, J., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does it tell us about speech production units? *Journal of Speech Language and Hearing Research*, 61(7), 1613–1625.
- Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2), 1–24. <http://doi.org/10.18637/jss.v034.i02>.
- Cristia, A., & Hallé, P. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.). *Speech planning and dynamics* (pp. 11–38). Frankfurt am Main: Peter Lang.
- Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In S. A. Brady, & D. P. Shankweiler (Eds.). *Phonological processes in literacy* (pp. 97–117). Hillsdale, NJ: Erlbaum.
- Goswami, U. (2006). Sensorimotor impairments in dyslexia: Getting the beat. *Human Learning*, 256–257. <https://doi.org/10.1111/j.1467-7687.2006.00484.x>.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>.
- Lefebvre, P., & Sutton, A. (2008). Phonological awareness tasks for French-speaking preschoolers. *Canadian Journal of Speech-Language Pathology and Audiology*, 32(4), 158–168.
- Liu, C., Cripe, T. P., & Kim, M. (2009). Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. *Molecular Therapy*, 18(9), 1724–1730. <https://doi.org/10.1038/mt.2010.127>.
- MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology*, 22, 113–117. <https://doi.org/10.1016/j.cub.2011.11.052>.
- McQueen, J. M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech*, 49(1), 101–112. <https://doi.org/10.1177/00238309060490010601>.
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352. <https://doi.org/10.1037/a0026744>.
- Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken word recognition. *Journal of Memory and Language*, 98, 77–92. <https://doi.org/10.1016/j.jml.2017.09.005>.
- Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2), 356–361. <https://doi.org/10.1016/j.cognition.2013.07.011>.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7(4), 323–331. [https://doi.org/10.1016/0010-0277\(79\)90020-9](https://doi.org/10.1016/0010-0277(79)90020-9).
- Mousty, P., Leybart, J., Alegria, J., Content, A., & Morais, J. (1994). BELEC: Une batterie d'évaluation du langage écrit et de ses troubles. In J. Grégoire, & B. Piérart (Eds.). *Évaluer les troubles de la lecture: Les nouveaux modèles théoriques et leurs implications diagnostiques* (pp. 127–145). Bruxelles: De Boeck.
- Nitrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language and Hearing Research*, 32(1), 120–132.
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>.
- Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-analysis. *Scientific Studies of Reading*, 19(5), 340–359. <https://doi.org/10.1080/10888438.2015.1052455>.
- Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience*, 30(25), 8435–8444. <https://doi.org/10.1523/JNEUROSCI.5791-09.2010>.
- R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria, 0, (ISBN 3-900051-07-0). <http://doi.org/10.1038/sj.hdy.6800737>.
- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717. <https://doi.org/10.1152/jn.00773.2011>.
- Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. Retrieved from In B. Lindblom, & S. E. G. Öhman (Eds.). *Frontiers of speech communication research* (pp. 217–229). New York: Academic Press. <http://search.ebscohost.com/login.aspx?direct=true&db=mzh&AN=1983016885&lang=fr&site=eds-live>.
- Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336–354. <https://doi.org/10.1016/j.jneuroling.2009.12.004>.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2).
- Shiller, D. M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), <https://doi.org/10.1037/a0036660>.
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343–348. [https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6).
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71. <https://doi.org/10.1037/1082-989X.11.1.54>.
- Stoodley, C. J., Harrison, E. P. D., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, 44(5), 795–798. <https://doi.org/10.1016/j.j>

- neuropsychologia.2005.07.009.
- Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, 66(1), 18–41. <https://doi.org/10.1006/jbrln.1997.1855>.
- Terband, H., & van Brenk, F. (2015). Compensatory and adaptive responses to real-time formant shifts in adults and children. *Proceedings of the 18th congress of phonetic sciences* (pp. 1–5). .
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>.
- Van Den Bunt, M. R., Groen, M. A., Frost, S., Lau, A., Preston, J. L., Gracco, V. L., ... Verhoeven, L. T. W. (2018). Sensorimotor control of speech and children's reading ability. *Scientific Studies of Reading*, 22(6), 1–14. <https://doi.org/10.1080/10888438.2018.1491583>.
- Van Den Bunt, M. R., Groen, M. A., Ito, T., Francisco, A. A., Gracco, V. L., Pugh, K. R., & Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit. *Journal of Speech Language and Hearing Research*, 60(3), 654. https://doi.org/10.1044/2016_JSLHR-L16-0201.
- Van Den Bunt, M. R., Groen, M. A., van der Kleij, S. W., Noordenbos, M. W., Segers, E., Pugh, K. R., & Verhoeven, L. (2018). Deficient response to altered auditory feedback in dyslexia. *Developmental Neuropsychology*, 43(7), 622–641. <https://doi.org/10.1080/87565641.2018.1495723>.
- Ventura, P., Kolinsky, R., Fernandes, S., Querido, L., & Morais, J. (2007). Lexical restructuring in the absence of literacy. *Cognition*, 105(2), 334–361. <https://doi.org/10.1016/j.cognition.2006.10.002>.
- Viana, A. R., Razuk, M., de Freitas, P. B., & Barela, J. A. (2013). Sensorimotor integration in dyslexic children under different sensory stimulations. *PLoS ONE*, 8(8), e72719. <https://doi.org/10.1371/journal.pone.0072719>.
- Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, 45(4), 683–725. <https://doi.org/10.1515/LING.2007.021>.
- Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108(1), 1–27. <https://doi.org/10.1111/bjop.12207>.
- Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, 16(1), 5–20. <https://doi.org/10.1023/A:1021789804977>.
- Wang, J., Wong, A. W. K., Wang, S., & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports*, 7(1), 5815. <https://doi.org/10.1038/s41598-017-06186-z>.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>.

4.4 Résultats complémentaires de l'étude 3 chez l'enfant

Dans cette partie, nous présentons des résultats complémentaires, issus de la 2^{ème} étude de ce chapitre. Ces résultats n'ont pas été évoqués dans l'article, puisque celui-ci est sous un format court. Il nous semble néanmoins utile de les décrire, puisqu'ils peuvent apporter un éclairage supplémentaire sur nos résultats, qui sera commenté dans la discussion générale de cette thèse. D'une part, nous examinerons la relation établie dans l'article entre adaptation et conscience phonologique, en distinguant les sous-composantes de celle-ci. D'autre part, nous examinerons la contribution de l'unité mot dans la définition de l'articulation de la parole, en fonction du groupe d'âge. Pour évaluer cette contribution, nous contrasterons le transfert et le post-effet de l'apprentissage auditorimoteur.

4.4.1 Conscience phonologique et adaptation

Dans l'article, nous avons montré que les enfants qui se sont adaptés à la perturbation auditive avaient en moyenne un meilleur score de conscience phonologique que les autres enfants, quel que soit le groupe d'âge. Nous aimerions préciser cette relation observée entre adaptation auditorimotrice et conscience phonologique. Dans la conscience phonologique, on peut distinguer la conscience des rimes ou des syllabes, qui se développent avec le langage oral avant l'apprentissage de la lecture, et la conscience phonémique, qui émerge en étroite interaction avec l'apprentissage de la lecture. Nous souhaitons déterminer si l'adaptation auditorimotrice est davantage reliée à l'une ou à l'autre de ces deux sous-composantes de la conscience phonologique.

4.4.1.1 Méthode et résultats

Rappelons que le test de conscience phonologique est composé de 3 tâches pour les enfants de 4-5 ans, et 3 tâches supplémentaires soit un total de 6 tâches pour les enfants de 7-8 ans. Les scores pour chaque tâche sont rappelés dans la table 4.1. Au total, le test inclut une tâche de conscience des rimes (le jugement de rimes), deux tâches de conscience syllabique (la suppression et l'inversion de syllabes) et trois tâches de conscience phonémique (suppression de phonèmes dans des CVC et des CCV, et inversion de phonèmes).

Phonological Awareness scores (in % of correct responses)		Preschoolers		Schoolchildren	
		mean	sd	mean	sd
Part 1	Rhyme judgement	58	25	94	7
	Syllable deletion	25	29	94	10
	Phoneme deletion in CVC	2	4	80	20
	Part 1 subtotal	31	17	90	9
Part 2	Phoneme deletion in CCV	/	/	78	21
	Syllable inversion	/	/	75	21
	Phoneme inversion	/	/	86	9
	Total score	31	17	84	10

Table 4-1: Phonological awareness scores in % of correct responses by age group and by subtask (extraite de Caudrelier et al., 2019, section 3.3)

Pour savoir si certaines tâches de conscience phonologique sont plus affectées que d'autres par l'adaptation, nous avons utilisé la même méthode que dans l'article, appliquée aux scores partiels de conscience phonémique d'une part, et de conscience de la syllabe et de la rime d'autre part (voir section 2.6.2 de l'article). Ainsi, les scores partiels, exprimés en ratio de réponses correctes, ont été modélisés avec les variables *age_group* (groupe d'âge) et *adaptation* (adaptés vs. non adaptés) à l'aide d'une régression Beta, réalisée avec la fonction *Betareg* de R (Cribari-Neto & Zeileis, 2010).

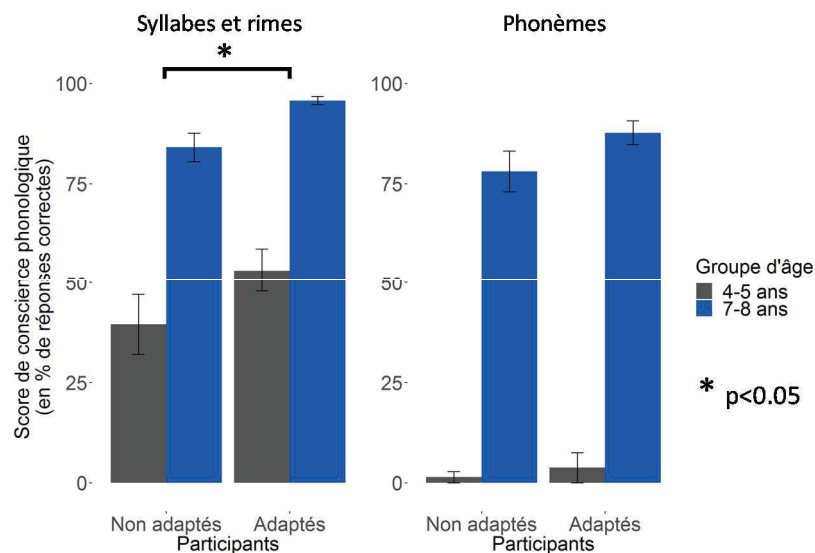


Figure 4.4: Scores partiels aux tâches de conscience phonologique relatifs à (1) la syllabe et la rime et (2) le phonème, en fonction du groupe d'âge et de si les enfants sont adaptés ou non.

Dans tous les cas, le modèle montrait une influence significative du groupe d'âge, les résultats des enfants de 7-8 ans (en proportion de réponses correctes) étant toujours plus élevés que ceux des enfants de 4 ans. Le score de conscience phonémique n'était pas significativement différent entre les enfants adaptés et non adaptés, bien que légèrement plus élevé chez les adaptés ($\chi^2(1)=1.16, p=0.3$). En revanche, les enfants adaptés, quel que soit leur groupe d'âge, ont obtenu un meilleur score de conscience de la syllabe et des rimes que les autres ($z=2.44, p=0.03$). De plus, chez les enfants de 8 ans, ceux qui se sont adaptés ont un meilleur score de conscience syllabique (soustraction et inversion syllabique) que les autres ($z=3.50, p=0.06$).

4.4.1.1 Discussion

Les résultats de cette analyse montrent que l'adaptation auditorimotrice est en lien avec la conscience de la syllabe et des rimes, et laissent penser que ce n'est pas le cas – ou dans une moindre mesure – pour la conscience phonémique. L'adaptation auditorimotrice permettant de sonder les représentations sensorimotrices, ce résultat suggère que la conscience de la syllabe et des rimes serait davantage reliée aux représentations sensorimotrices que la conscience phonémique.

En réalité, chez les 4-5 ans, le faible lien entre adaptation et conscience phonémique peut s'expliquer simplement par le fait que la tâche de conscience du phonème est peu informative, puisque la vaste majorité des enfants ont obtenu un score de 0 (seuls 3 enfants ont eu un score différent de 0)¹. Néanmoins, le résultat est valable chez les 7-8 ans, indépendamment des 4-5 ans : les enfants adaptés ont eu de meilleurs scores en conscience syllabique que les autres, mais pas en conscience phonémique, du moins pas à un niveau significatif, alors qu'il n'y avait pas d'effet plafond ni plancher sur les scores.

Nos observations suggèrent que certains enfants de 7-8 ans utilisaient une stratégie orthographique pour répondre aux tâches de conscience phonémique (Stuart, 1990). Par exemple, pour inverser /az/, certains enfants répondaient /ga/ au lieu de /za/. Notre interprétation est qu'ils visualisaient probablement le mot « âge » (qui se prononce /az/, inversaient les lettres

¹ Ce score n'est pas étonnant, puisque l'on sait que la conscience phonémique se développe en général simultanément avec l'apprentissage de la lecture. Ainsi, le score de conscience phonémique chez l'enfant de 4-5 ans était destiné à contrôler que ces enfants n'étaient pas encore capables de lire, et que leur score était significativement différent du groupe de lecteurs (7-8 ans). Par conséquent, chez les 4-5 ans, seules les deux épreuves de conscience de la syllabe et des rimes ont permis d'obtenir des scores avec une variabilité interindividuelle importante.

non muettes, pour former 'ga', et répondaient ainsi /ga/. Ainsi, l'utilisation d'une stratégie orthographique a pu influencer les scores de conscience phonémique. Il est possible que cette influence amenuise les liens entre la mesure de la conscience phonémique et la qualité des représentations sensorimotrices. Ainsi, les scores de conscience syllabique traduisent probablement mieux la qualité des représentations phonologiques des enfants que les scores de conscience phonémique.

D'autre part, on peut se demander à quel(s) niveau(x) de granularité en termes de représentations phonologiques se situe le lien conscience phonologique – adaptation. Plus précisément, l'adaptation auditorimotrice traduit-elle une spécification au niveau phonémique ? Ou une spécification syllabique suffit-elle pour expliquer les phénomènes d'adaptation observés? Notre étude ne permet pas de répondre à la question. Cela renforce néanmoins le besoin d'étudier le transfert d'apprentissage auditorimoteur, afin de mieux distinguer les représentations sensorimotrices des niveaux syllabique et phonémique - ou infra-syllabique.

4.4.2 Ratio de transfert

Pour la production de ses premiers mots, le bébé semble utiliser des représentations holistiques, les templates (Vihman & Croft, 2007). Ensuite, des représentations de niveaux de granularité plus fine se développeraient graduellement (Walley, Metsala, & Garlock, 2003, voir aussi la section 1.4 Développement des unités phonologiques). Nous cherchions dans l'article à savoir si les représentations phonémiques se construisent avant ou simultanément à l'apprentissage de la lecture. Cette question est en effet matière à débat dans les théories du développement phonologique (Ainsworth, 2015; Ventura, Kolinsky, Fernandes, et al., 2007; Ziegler & Goswami, 2005).

Les résultats de l'étude révèlent un transfert d'apprentissage auditorimoteur au niveau du phonème, chez tous les groupes d'âge, suggérant l'émergence de représentations phonémiques en production, dès l'âge de 4-5 ans. On peut néanmoins se demander si ces représentations phonémiques ont le même poids dans la définition de l'articulation chez l'enfant non-lecteur de 4-5 ans que chez l'enfant lecteur et l'adulte. Il est possible que l'émergence du phonème se fasse de manière graduelle, et que dans un premier temps les représentations holistiques du mot (désigné par l'unité mot) gardent une place prépondérante dans le contrôle de la parole. Cette place pourrait ensuite diminuer avec l'expérience du langage oral, et avec l'apprentissage de la lecture.

Pour évaluer la contribution relative des représentations segmentales par rapport à l'unité mot, nous contrastons le transfert (aux niveaux syllabique et phonémique) avec le post-effet de l'apprentissage auditorimoteur. En effet, le post-effet désigne la persistance du changement de prononciation après l'arrêt de la perturbation auditive, sur le même mot que celui utilisé lors de l'entraînement. Le post-effet cumule donc les contributions à l'échelle du phonème, de la syllabe et du mot, pour le contrôle de la parole. Contraster post-effet et transfert permet ainsi de distinguer la contribution du mot.

4.4.2.1 Méthode et résultats

Un modèle linéaire mixte (LMM) unique a été bâti pour comparer transfert et post-effet, chez les sujets adaptés, avec pour facteurs fixes la phase (transfert vs post-effet), le groupe d'âge (4-5 ans, 7-8 ans et adultes), et l'essai (voir la section 2.6.1 de l'article pour des précisions sur la construction des LMM). Pour ce faire, le facteur *syllabe* (1^{ère} vs 2^{ème} syllabe de 'épée') a été exclu puisqu'il n'a pas d'effet dans le LMM dédié au transfert, présenté dans l'article, et qu'il n'y a qu'une seule syllabe dans le post-effet.

Le modèle obtenu est constitué de la somme de l'interaction entre la phase et le groupe d'âge ($\chi^2(2)=7.3$, $p=0.03$) et de l'interaction entre la phase et l'essai ($\chi^2(1)=5.35$, $p=0.02$) : $\text{amplitude} \sim \text{groupe_age} * \text{phase} + \text{essai} * \text{phase}$.

La figure 4.3 montre l'amplitude en début de phase (ou intercept) en fonction du groupe d'âge et de la phase. L'amplitude du post-effet est significativement plus élevée que celle du transfert chez les enfants de 4-5 ans ($+4.3\% \pm 1.0$; $z=4.1$; $p<0.001$) mais pas chez les 7-8 ans ($+0.8\% \pm 1.1$; $z=0.8$; $p=0.8$) ni chez les adultes ($+1.0\% \pm 0.9$; $z=1.1$; $p=0.7$). D'autre part, indépendamment du groupe d'âge, le transfert décroît légèrement plus rapidement que le post-effet, la différence n'étant pas tout à fait significative ($-0.08\%/\text{essai} \pm 0.03$; $z=2.3$; $p=0.07$).

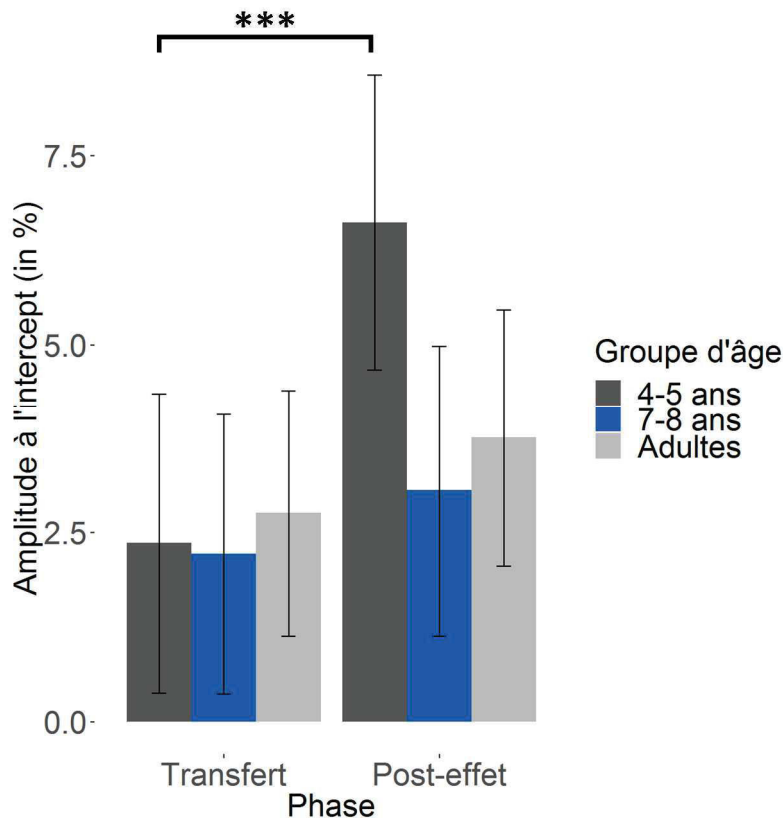


Figure 4.5: Amplitude en début de phase (intercept du modèle) par phase et par groupe d'âge. Les barres indiquent les intervalles de confiance à 95%. *** $p < 0.001$

4.4.2.1 Discussion

Emergence graduelle du phonème chez l'enfant

Nous avons utilisé la différence d'amplitude entre transfert et post-effet comme signe de la contribution relative de l'unité mot dans le contrôle de la parole. Nos résultats montrent que cette différence est significative seulement chez les 4-5 ans. Ce résultat suggère que les enfants de 4-5 ans se baseraient davantage sur l'unité mot (ou représentation holistique du mot) et utiliseraient relativement moins les représentations segmentales que les enfants de 7-8 ans et les adultes.

Ce résultat est cohérent avec la proposition de Vihman & Croft (2007) selon laquelle la première unité utilisée par le bébé est le mot lui-même, (ou plutôt un *template* de mot), avant l'émergence des représentations de niveaux de granularité plus fine. Il vient cependant nuancer les travaux indiquant l'apparition de représentations segmentales dès l'âge de 4 ans, et en particulier nos résultats présentés dans l'article.

On peut se demander si cette évolution est directement causée par l'apprentissage de la lecture, ou si elle accompagne le développement « naturel » de l'enfant, découlant de son expérience du langage oral. Nos résultats montrent que cette différence d'amplitude entre transfert et post-effet est tout à fait équivalente entre enfants de 7-8 ans et adultes. Cela pourrait suggérer que l'acquisition de la lecture et de l'écriture joue un rôle dans l'évolution de l'utilisation de représentations holistiques vers des représentations plus segmentales. Néanmoins, seule la comparaison de populations lettrées et illettrées (adultes ou du même âge) permettrait de trancher cette question. Si une différence était observée entre ces populations, cela permettrait de proposer une vision intermédiaire dans le débat sur le lien entre apparition de représentations phonémiques implicites et acquisition de la lecture : l'apparition des représentations phonémiques précéderait l'acquisition de la lecture, mais l'usage de la lecture viendrait renforcer le rôle des représentations phonémiques dans la production de parole.

Par ailleurs, deux points techniques importants concernant l'évaluation du post-effet peuvent être soulevés par ces résultats. Ils sont développés dans les paragraphes suivants.

Post-effet et sélection des participants adaptés

D'abord, il est à noter que seuls les sujets adaptés ont été sélectionnés dans cette analyse du post-effet, en cohérence avec le transfert, qui ne peut avoir lieu que s'il y a eu adaptation au préalable. A l'inverse, dans l'article (Caudrelier et al., 2019), l'analyse conjointe de l'adaptation et du post-effet a été menée sur tous les participants. Rappelons que lorsque tous les participants sont pris en compte, l'amplitude du post-effet, de même que celle de l'adaptation, est équivalente entre les groupes d'âge. Au contraire, lorsque l'on sélectionne uniquement les participants adaptés, l'amplitude de l'adaptation et du post-effet est plus élevée chez les enfants de 4 ans que les autres. Cela est probablement lié à leur variabilité de production. Comme les sujets adaptés sont sélectionnés selon un critère de différence statistique entre la baseline et la phase d'adaptation, il est nécessaire d'atteindre une amplitude d'adaptation plus importante pour qu'elle soit significativement plus grande que 0, malgré la variabilité des productions. Cela explique d'ailleurs peut-être pourquoi le taux de sujets adaptés est plus bas chez l'enfant que chez l'adulte. Utiliser une valeur seuil d'amplitude comme critère d'adaptation aurait sans doute permis de contourner cette limitation, mais poserait d'autres problèmes, en particulier celui du caractère arbitraire du choix du seuil. Notons que ce point ne remet en cause aucun de nos deux résultats principaux concernant l'émergence du phonème dès l'âge de 4-5 ans, et se poursuivant jusqu'à 7-8 ans.

Peut-on comparer transfert et post-effet alors qu'ils se succèdent?

On peut se demander s'il est possible de comparer le transfert et le post-effet alors que l'un a lieu après l'autre, et qu'un potentiel effet d'ordre n'est pas contrôlé. Pour être parfaitement rigoureux, il aurait probablement fallu constituer un autre groupe suivant une autre condition avec post-effet juste après la phase d'adaptation, sans transfert au préalable, comme dans notre étude précédente (Caudrelier, Schwartz, et al., 2018). L'étude précédente en question nous a montré que les effets de la phase de transfert sur l'amplitude du post-effet sont très limités, et ne concernaient que la pente : il n'y avait pas de différence entre conditions sur l'intercept du post-effet. Par ailleurs, on pourrait s'attendre à ce que l'apprentissage disparaisse en partie pendant la phase de transfert, et donc que le post-effet démarre plus bas que s'il avait suivi directement la phase d'adaptation. Or, dans notre étude le post-effet a le même intercept que l'adaptation, ce qui laisse penser qu'il repart exactement au même niveau que s'il n'y avait pas eu de phase de transfert entre les deux. Et même si le post-effet ne repartait pas au même niveau que l'adaptation, deux cas sont possibles. (1) Soit la phase de transfert aurait affecté indifféremment les 3 groupes d'âge, auquel cas cela ne changerait rien au résultat : l'écart entre transfert et post-effet est plus élevé chez les 4-5 ans que chez les autres. (2) Soit la phase de transfert aurait davantage fait diminuer le post-effet chez les 7-8 ans et les adultes que chez les enfants de 4-5 ans. Le fait que post-effet et transfert soient davantage reliés entre eux indique que le post-effet repose moins sur un apprentissage spécifique au mot, que sur les représentations sublexicales communes avec celles du transfert. Ce raisonnement suggérerait donc aussi que les 7-8 ans et les adultes utilisent davantage les représentations sublexicales que les 4-5 ans, ces derniers donnant l'avantage aux représentations du mot.

Ainsi, après avoir analysé tous les cas possibles, nous confirmons que notre résultat est valide : la comparaison du transfert et du post-effet selon le groupe d'âge indique que l'importance de la contribution de l'unité mot décroît entre 4-5 ans et 7-8 ans, au profit des représentations sublexicales (syllabes et phonèmes).

4.5 Conclusion

Ces deux études utilisent le paradigme de transfert d'apprentissage auditorimoteur pour sonder les représentations sensorimotrices à la base des interactions entre parole, lecture et conscience phonologique. Tandis que la première suggère une influence de la lecture sur l'activation des représentations sensorimotrices lors de la tâche, à travers l'effet de la modalité de présentation du stimulus, la deuxième suggère des capacités de transfert phonémique commune entre le groupe d'enfants pré-lecteurs et les groupes de lecteurs, mais aussi sans doute un rôle plus important du mot en tant qu'unité dans le contrôle de la parole chez les pré-lecteurs que chez les lecteurs. Elle montre aussi un lien entre adaptation et conscience phonologique, que nous interprétons comme un rôle de précurseur joué par les représentations sensorimotrices sur la conscience phonologique.

En reprenant une distinction classique entre processus de hauts-niveaux, comme la lecture et la conscience phonologique, et processus sensorimoteurs, dits de bas-niveaux, nos résultats suggèrent un effet *top-down* (du haut vers le bas) immédiat d'une part, et d'autre part des effets structurants, de type *bottom-up* (de bas en haut), sur une échelle de temps plus longue, dans laquelle s'inscrit le développement de l'enfant.

5. Discussion générale

Les objectifs principaux de cette thèse étaient de déterminer le niveau de granularité (mot, syllabe, phonème) des représentations qui sous-tendent la production de parole à l'âge adulte, et au cours du développement de l'enfant. Cette question de l'unité, très présente dans la littérature de psycholinguistique, a été traitée ici avec une méthode expérimentale issue de la recherche sur le contrôle moteur : la perturbation du retour auditif lors de la production de la parole, et l'observation de l'adaptation qui en résulte, dite adaptation auditorimotrice, et son transfert. L'exploitation de cette méthode dans le cadre de l'étude des unités de la parole s'appuie sur une hypothèse centrale de cette thèse, celle de la nature sensorimotrice des unités de parole. Après un rappel des principaux résultats expérimentaux, nous proposons une réflexion sur la portée de ces résultats, leurs limitations potentielles, et les perspectives de recherche associées, autour de 4 grands thèmes ou questions, faisant chacun l'objet d'une section.

5.1. Rappel des résultats principaux

Les chapitres 3 et 4 ont été consacrés à la présentation des trois principales études expérimentales réalisées dans le cadre de cette thèse afin : (1) de contraster le rôle du phonème, de la syllabe et du mot dans la production de la parole chez l'adulte ; d'évaluer les effets de la lecture sur l'unité de production (2) en étudiant l'effet de la modalité (mot écrit vs image) de présentation chez l'adulte et (3) en estimant le rôle des différentes unités chez l'enfant, en lien avec le développement de la conscience phonologique, dans une étude transversale. Une synthèse des paramètres contrôlés ayant varié d'une étude à l'autre est présentée en table 5.1.

Table 5-1: Synthèse des paramètres ayant varié dans les trois études présentées dans cette thèse. Le paramètre dont l'influence est évaluée dans l'étude est mis en gras.

	Chapitre 3		Chapitre 4
Etude (lieu de publication)	1 (JSLHR)	2 (Interspeech)	3 (Cognition)
Tranches d'âge	Adultes	Adultes	Adultes vs enfants
Lieu de recrutement	France	France	Québec
Modalité de présentation	Mot écrit	Mot écrit vs image	Image
Stimulus d'entraînement	bé	pé (/pe/)	/pe/
Stimulus de test (ou de transfert)	bé, bépé, pébé, bébé	épée	épée
Durée de l'entraînement	180 essais	180 essais	50 essais

5.1.1 Un transfert d'apprentissage pour l'analyse des rôles du phonème, de la syllabe et du mot

Une première originalité de notre travail a consisté à utiliser le paradigme de transfert d'apprentissage avec des séquences de plusieurs syllabes, porteuses de sens ou pas. En effet, ce paradigme avait jusqu'alors été exploité uniquement dans des monosyllabes, rendant impossible toute distinction entre l'unité syllabe et l'unité mot. Notre étude chez l'adulte (Chapitre 3) a été conçue pour contraster le transfert selon des niveaux de granularité différents en termes d'unités linguistiques : phonème, syllabe et mot. Elle a ainsi utilisé le transfert pour mieux évaluer la contribution respective de ces unités dans l'articulation de séquences de parole. Après une adaptation sur le pseudo-mot /be/, nous avons mesuré le post-effet sur ce même pseudo-mot /be/, ainsi que le transfert de la syllabe /be/ vers les pseudo-mots /bepe/ et /pebe/ et vers le mot /bebe/ ('bébé') - désignés ci-après par le terme d'énoncés de test. Nous avons obtenu les résultats suivants :

(1a) le transfert était significatif vers toutes les syllabes de tous les énoncés tests, qui contenaient la même voyelle /e/. Il y a donc eu transfert au niveau du phonème /e/, ou du moins à un niveau infra-syllabique, que l'on désignera dans la suite de cette discussion par l'*effet phonème*.

(1b) le transfert vers la syllabe /be/ était significativement plus élevé que le transfert vers la syllabe /pe/, lorsque celles-ci étaient en première position des séquences bisyllabiques. Il y a donc eu, en plus du transfert sur le phonème, un transfert spécifique au niveau de la syllabe, ou *effet syllabe*.

(1c) le transfert vers les syllabes /be/ dans les trois énoncés de tests était d'amplitude plus faible que le post-effet sur l'énoncé d'entraînement /be/. Plus spécifiquement, la dynamique du transfert au cours des répétitions diffère de celle du post-effet, et une discontinuité a été observée pour la prononciation de /be/ entre les blocs de transfert (avec les énoncés de test) et les blocs de post-effet (avec l'énoncé d'entraînement). Il y a donc eu un effet de l'énoncé lui-même, au-delà de la syllabe, que l'on désignera pour simplifier, par abus de langage, par *effet mot*.

(1d) enfin, pour la syllabe /be/, l'amplitude du transfert était plus importante quand elle était placée en première position des items bisyllabiques qu'en 2^{ème} position (ci-après, *effet de position*). L'influence de la place dans le mot conforte aussi l'observation de l'influence d'une unité qui dépasse la taille de la syllabe, et rejoint notre conclusion ci-dessus sur le rôle du mot (1c).

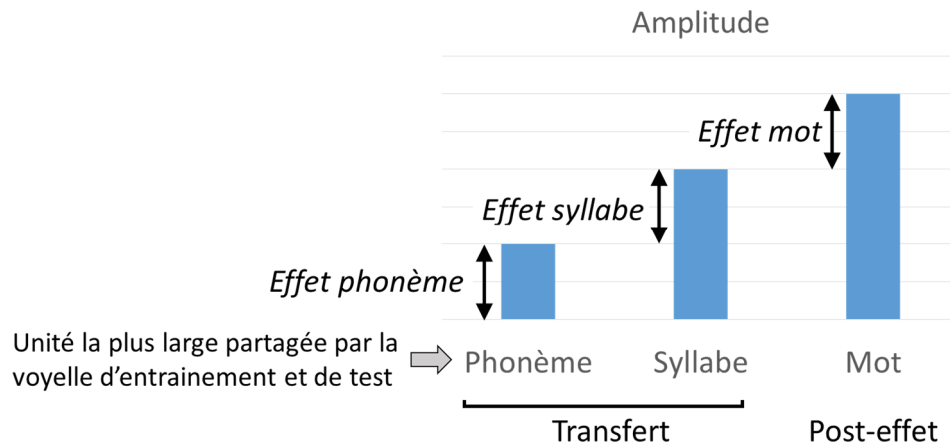


Figure 5-1: Représentation schématique des effets phonème, syllabe et mot, en relation avec les amplitudes de transfert et de post-effet

La Figure 5.1 représente qualitativement les effets phonème, syllabe et mot, en relation avec les amplitudes de transfert et de post-effet. Comme nous l'avons mentionné ci-dessus, les études expérimentales précédentes n'avaient pas permis de séparer l'effet syllabe de l'effet mot. En effet, il avait simplement été observé que le post-effet sur un énoncé monosyllabique de type CVC était d'amplitude supérieure au transfert sur un énoncé du même type, contenant des consonnes et/ou une voyelle différentes (Rochet-Capellan, Richer, & Ostry, 2012). En mesurant ces différents effets dans une même étude expérimentale, nous avons mis en évidence l'influence de plusieurs niveaux d'unités linguistiques dans l'articulation de séquences de parole.

5.1.2 L'influence de la modalité dans le transfert d'apprentissage

L'étude préliminaire du chapitre 4 contrastait, chez l'adulte, l'adaptation sur /pe/ et le transfert à /epe/ selon la modalité de présentation du stimulus : mot à lire ou image à dénommer. Nous avons observé :

(2a) une adaptation et un post-effet significatifs et équivalents quel que soit le mode de présentation ;

(2b) pour le transfert, une interaction entre syllabe et modalité de présentation du stimulus de transfert. En effet, le transfert au /pe/ de /epe/ était plus grand que le transfert au premier /e/ de /epe/, seulement lorsque le stimulus était un mot à lire, mais pas quand il s'agissait d'une image à nommer. Ainsi, on peut dire que l'effet syllabe (décrit en 1b) était présent seulement avec le mot écrit, mais pas avec l'image. Cette observation sera désignée par l'effet de modalité.

En outre, l'analyse comparative d'une partie des données issues des études 2 et 3, détaillée en Annexe 1, suggère que :

(2c) chez l'adulte, le post-effet a une amplitude significativement plus élevée que le transfert seulement lorsque la phase d'entraînement (avec perturbation auditive) est longue (180 essais) mais pas lorsqu'elle est courte (50 essais). Ainsi, l'*effet mot* dépendrait de la durée de l'entraînement.

5.1.3 Les représentations phonologiques révélées par le transfert chez l'enfant

L'étude principale présentée dans le chapitre 4 est la première étude de transfert d'apprentissage auditorimoteur chez l'enfant, à notre connaissance. Après avoir mis en évidence la contribution de plusieurs unités dans l'articulation de la parole chez l'adulte, nous nous sommes intéressés au développement de ces unités chez l'enfant. Dans la littérature, s'il semble y avoir consensus sur le fait que le mot et la syllabe se développent dans les deux premières années de vie, la construction de la représentation des phonèmes est davantage sujette à débat. En particulier, la lecture pourrait avoir une influence sur celle-ci. Nous nous sommes donc demandés si la représentation sensorimotrice du phonème apparaissait avant l'apprentissage de la lecture ou simultanément à celui-ci.

Après une adaptation sur le pseudo-mot /pe/, le transfert au mot /epe/ (épée) a été testé chez des enfants en âge préscolaire, c'est-à-dire avant apprentissage de la lecture (4-5 ans), en âge scolaire, soit après apprentissage de la lecture (7-8 ans), et chez des adultes. Nous avons observé dans cette étude :

(3a) un transfert significatif sur les deux syllabes de /epe/ chez tous les groupes d'âge. En particulier, le transfert significatif au premier /e/ de /epe/ montre un transfert sur la voyelle indépendamment de la structure syllabique. Nous avons ainsi trouvé l'*effet phonème* chez les enfants, y compris avant apprentissage de la lecture. Ce résultat soutient l'idée que le phonème se développe indépendamment de l'apprentissage de la lecture, simplement avec le développement du langage oral.

(3b) l'amplitude du post-effet est significativement plus élevée que celle du transfert chez les enfants de 4-5 ans, mais pas chez les 7-8 ans ni les adultes. Cet écart entre transfert et post-effet est le signe d'un *effet mot*. Ce résultat suggère que pour la production de parole, l'enfant d'âge préscolaire utiliserait davantage les représentations du mot et moins les représentations segmentales que les deux autres groupes d'âge.

(3c) un lien entre adaptation auditorimotrice et conscience phonologique. En effet, dans chacun des deux groupes d'âges, les enfants qui montraient une adaptation significative à la perturbation auditive avaient en moyenne une conscience phonologique plus élevée que ceux qui ne s'adaptaient pas. Ce point fait l'objet d'une discussion plus spécifique dans la section suivante.

5.2. Relations entre adaptation auditorimotrice et conscience phonologique

5.2.1 Une relation de cause à effet ?

Dans notre étude chez l'enfant (Caudrelier et al., 2019), un lien a été mis en évidence entre adaptation auditorimotrice, et conscience phonologique. Plus précisément, nous l'avons dit ci-dessus, les enfants qui se sont adaptés significativement, en modifiant leurs formants dans la direction opposée à la perturbation, avaient une meilleure conscience phonologique que les autres. Cela était vrai dans les deux groupes d'âge. Ces résultats étendent ceux de l'étude de Van Den Bunt et al. (2018), dans laquelle une corrélation était observée entre amplitude d'adaptation et niveau de conscience phonologique chez les enfants d'âge préscolaire uniquement (4-5 ans). Une réflexion sur la nature de ce lien a été brièvement entamée dans l'article du chapitre 4 (étude chez l'enfant), réflexion que nous allons développer ici.

Une première interprétation possible est que le développement de la conscience phonologique influencerait l'adaptation auditorimotrice. Si c'était le cas, les enfants de 7-8 ans, qui ont en moyenne une bien meilleure conscience phonologique que les enfants de 4-5 ans, devraient s'adapter à la perturbation en plus grande proportion, et/ou selon une amplitude moyenne supérieure. Or, les enfants de 7-8 ans ne montrent en moyenne pas plus d'adaptation que les enfants de 4-5 ans, que ce soit au regard du pourcentage d'enfant adaptés, ou de l'amplitude de l'adaptation. Cette interprétation n'est donc pas cohérente avec nos données.

Une autre possibilité est que l'adaptation auditorimotrice serait un prédicteur de la conscience phonologique à un âge ultérieur, puisque l'adaptation reste stable entre les groupes d'âge alors que la conscience phonologique s'améliore. Il n'est pas exclu que les enfants du groupe adaptés à 4-5 ans ne s'adaptent plus à 7-8 ans et inversement. Seule une étude longitudinale permettrait de le tester. Néanmoins cela semble peu probable, étant donné que la conscience phonologique chez les enfants d'âge préscolaire est un prédicteur de leur conscience phonologique à l'âge de 7-8 ans (Anthony & Lonigan, 2004; Carroll, 2001). La Figure 5.2 représente un modèle

d'évolution où l'adaptation resterait stable entre les deux tranches d'âge, et permettrait de prédire la conscience phonologique ultérieure.

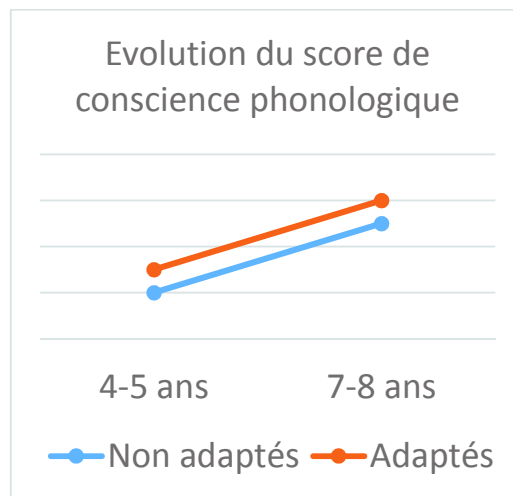


Figure 5-2 : Modèle qualitatif représentant le cas où l'adaptation serait stable au cours du développement, et constituerait ainsi un prédicteur de la conscience phonologique

Ce possible rôle de prédicteur de l'adaptation auditorimotrice sur la conscience phonologique ne signifie pas nécessairement qu'il y a un lien de cause à effet entre les deux. Le lien observé pourrait être lié à l'influence d'un autre facteur sur ces deux compétences, comme évoqué dans l'article. Il est souvent affirmé que la conscience phonologique repose sur la qualité des représentations phonologiques. Mais ce terme de qualité est difficile à définir précisément, à quantifier. Par exemple, dans Carroll (2001), la qualité des représentations phonologiques est évaluée à travers la prononciation correcte de mots. La qualité des représentations pourrait également être définie par la netteté des frontières catégorielles, qui influencent à la fois la conscience phonologique et l'adaptation auditorimotrice (Lametti, Krol, Shiller, & Ostry, 2014; Niziolek & Guenther, 2013; Shiller & Rochon, 2014). Néanmoins ces mesures font intervenir soit la production, soit la perception, mais pas l'interaction perception-action. Une autre mesure possible est la performance lors d'une tâche de répétition de pseudo-mots. Dans une étude comparant l'adaptation auditorimotrice chez des enfants de 4 à 8 ans avec une apraxie développementale de la parole et chez un groupe d'enfants contrôle, Terband, van Brenk, & van Doornik-van der Zee (2014) ont montré une forte corrélation entre performance dans la répétition de pseudo-mots et amplitude d'adaptation. Ils ont aussi montré que les enfants avec apraxie ne compensaient pas la perturbation auditive, mais avaient au contraire tendance à aller dans le même sens qu'elle (ce qui résulte en un retour auditif encore plus éloigné du son

demandé). Cet ensemble d'observations suggère que l'adaptation auditorimotrice, en tant que mesure de l'intégration sensorimotrice, est un marqueur de la qualité des représentations phonologiques.

Une hypothèse serait que l'intégration sensorimotrice, telle que mesurée avec l'adaptation auditorimotrice joue un rôle de médiateur entre les compétences perceptives ou articulatoires seules, et les compétences reliées au développement phonologique de plus haut niveau, comme la conscience phonologique, ou même la mémoire phonologique telle que mesurée avec la répétition de pseudo-mots. Une meilleure compréhension de ces liens pourrait avoir des conséquences pour le diagnostic et la prise en charge de troubles développementaux comme la dyslexie ou l'apraxie de la parole. Certaines études longitudinales intégrant de nombreux facteurs utilisent des modèles statistiques pour déterminer entre autres les facteurs qui jouent un rôle de médiateur entre deux compétences (Hulme et al., 2012). Une telle étude longitudinale englobant des mesures d'adaptation sensorimotrice permettrait de tester cette hypothèse d'un rôle de médiateur joué par l'intégration sensorimotrice. Un tel rôle offrirait une base pour appuyer l'idée selon laquelle l'intégration sensorimotrice est ce qui permet l'émergence de catégories et de concepts. Cette idée est présente dans certains cadres théoriques évoqués précédemment, notamment les évolutions récentes de la théorie de l'exemplaire et des modèles multi traces de la mémoire (Versace et al., 2009).

5.2.2 Niveau syllabique ou phonémique ?

Le lien entre adaptation et conscience phonologique peut être précisé selon les sous-composantes de celle-ci. Si l'on rentre dans le détail des tests de conscience phonologique de notre étude (voir la partie « Résultats complémentaires » du chapitre 4), on observe que ce lien est plus marqué pour la conscience syllabique que la conscience phonémique, que ce soit chez les 4-5 ans ou chez les 7-8 ans. Cela peut suggérer que la conscience syllabique est davantage ancrée dans les représentations sensorimotrices, tandis que la conscience phonémique est aussi sous l'influence d'autres phénomènes, notamment de stratégies orthographiques.

Cette observation d'un lien avec la conscience syllabique peut conduire à une autre question : est-ce que l'adaptation auditorimotrice n'intervient que si les représentations lexicales du locuteur sont définies au niveau phonémique? Ou est-ce qu'un niveau de granularité syllabique suffit pour observer de l'adaptation ? Cela rejoint les divergences d'interprétation des travaux d'Eimas et al. (1971) sur la discrimination de syllabes chez le bébé. La capacité à discriminer

deux syllabes ne différant que par un seul phonème a d'abord été interprétée comme de la discrimination phonémique. Dans la mesure où la perturbation de formants est appliquée à une voyelle dans un contexte syllabique (/be/ dans la première étude, /pe/ dans les autres) nous pensons que l'existence de représentations syllabiques suffit à expliquer le phénomène d'adaptation.

Cette réflexion met en exergue l'utilité d'étudier le transfert, et pas seulement l'adaptation, pour mieux distinguer les contributions des niveaux syllabique et phonémique (ou infra-syllabique) dans l'articulation de la parole. La suite de cette discussion est consacrée aux résultats sur le transfert d'apprentissage.

5.3. L'effet « phonème »

5.3.1 Phonème, allophone, rime, trait phonétique et/ou geste articulatoire ?

Dans l'étude 1 chez l'adulte (chapitre 3) nous avons observé du transfert d'apprentissage auditorimoteur de la monosyllabe /be/ vers la syllabe /pe/ dans les items bisyllabiques /bepe/ et /pebe/. Les syllabes /be/ et /pe/ partageant le phonème /e/, nous en avons déduit que le transfert avait lieu au niveau du phonème /e/, ou en tout cas à un niveau infra-syllabique. Nous l'avons ainsi désigné comme l'*effet phonème*. Néanmoins, les syllabes /be/ et /pe/ partagent potentiellement d'autres unités, ou caractéristiques phonétiques, telles que le trait phonétique, le geste articulatoire, l'allophone ou la rime (puisque'il s'agit d'une syllabe CV). Dans notre étude, nous n'avons pas cherché à contraster l'amplitude du transfert sur ces différents niveaux, puisque nous nous sommes concentrés sur d'autres unités : le mot (ou énoncé), et la syllabe. Aussi, le transfert de /be/ à /pe/ pourrait se situer indifféremment au niveau de ces unités partagées.

Dans la phonologie articulatoire, l'unité de base est le geste articulatoire. Ainsi, les séquences /be/ et /pe/ partagent les gestes (i) de fermeture des lèvres par l'action conjointe de la mandibule et des lèvres, pour réaliser ces deux consonnes occlusives bilabiales, (ii) de constriction antérieure de la langue pour la réalisation du /e/, mais (iii) différent par la coordination entre constriction supra-glottiques et geste glottique différenciant la consonne voisée /b/ et la non-voisée /p/. L'adaptation à la perturbation que nous avons appliquée sur F1 et F2 devrait affecter les gestes de la mandibule et/ou de la langue mais pas directement ceux des cordes vocales. Donc on peut supposer que si le transfert avait lieu uniquement au niveau du geste articulatoire,

il devrait être sensiblement équivalent sur les syllabes /be/ et /pe/ dans les énoncés de test. Or ce n'est pas le cas. Il y a clairement plus de transfert sur /be/ que sur /pe/. C'est ce que nous avons appelé *l'effet syllabe*. Ainsi, l'hypothèse du transfert limité au niveau du geste articulatoire ne nous paraît pas convenir pour interpréter l'ensemble de nos résultats. Mais le geste articulatoire pourrait tout de même apporter sa contribution dans *l'effet phonème*.

La rime est aussi une unité considérée, en particulier dans le développement phonologique de l'enfant, notamment dans les travaux sur la conscience phonologique (Ziegler & Goswami, 2005). Pour faire la distinction entre transfert au niveau du phonème et de la rime, nous aurions pu utiliser des syllabes CVC, plutôt que des CV. Néanmoins, il nous paraissait important d'utiliser des syllabes CV, beaucoup plus courantes en français que des CVC (Rousset, 2008). Pour cette raison, l'unité rime nous paraît d'ailleurs moins pertinente en français qu'elle ne l'est par exemple en anglais, où les syllabes CVC sont très répandues. Notre choix de corpus est donc tout à fait justifié, mais il pourrait être complété dans de futurs travaux.

En perception, plusieurs études sur le transfert d'apprentissage, citées dans le chapitre 1 de cette thèse, soutiennent l'idée que l'allophone est une unité de base de la perception, et non pas le phonème (Mitterer et al., 2018, 2013). Dans nos études, l'allophone de /e/ était le même pour les mots d'entraînement et de test. Notre étude ne permet donc pas de tester si l'allophone joue un rôle.

Enfin, le transfert désigné comme *l'effet phonème* pourrait en fait traduire une proximité acoustique dans un espace continu. Cet espace continu pourrait par exemple être l'espace de sortie acoustique des modèles internes directs². Des travaux en anglais avec des mots d'entraînement et de test monosyllabiques de type CVC ont d'ailleurs montré que l'amplitude du transfert dépendait de la proximité acoustique entre les voyelles d'entraînement et de test (Rochet-Capellan et al., 2012). Cependant, dans notre étude chez l'adulte, la proximité acoustique telle que nous l'avons mesurée en milieu de voyelle dans l'espace formantique (F1,F2), ne permet pas de rendre compte des effets de transfert observés. La proximité acoustique pourrait néanmoins constituer un facteur parmi d'autres, qui gagnerait d'ailleurs à être redéfini, en y intégrant par exemple la fréquence fondamentale F0. F0 pourrait en effet jouer un rôle dans notre étude dans la mesure où les stimuli de test étaient des items

² Les modèles internes directs (*forward internal models*), présents dans la plupart des modèles de contrôle de la parole, mettent en relation commandes motrices et conséquences sensorielles de l'articulation de la parole.

bisyllabiques, au cours desquels la fréquence fondamentale peut varier beaucoup. L'évaluation de la proximité acoustique pourrait également être étendue à la syllabe sur toute sa durée, et pas seulement en milieu de voyelle.

Ainsi, le transfert à un niveau infra-syllabique gagnerait à être précisé. Pour cela, il serait judicieux de mener une autre étude sur le transfert d'apprentissage auditorimoteur, en contrastant cette fois les niveaux cités dans ce paragraphe : phonème, allophone, rime et geste articulatoire. L'utilisation de plusieurs structures syllabiques (CV, CVC), toujours dans des énoncés pluri-syllabiques, serait appropriée. En outre, étudier le transfert en fonction de la prosodie, notamment la fréquence fondamentale, permettrait de pousser les investigations concernant un possible effet de la proximité acoustique. Néanmoins, l'effet observé indique qu'au moins un élément au niveau infra-syllabique contribue à l'articulation de la parole.

5.3.2 L'émergence du phonème chez l'enfant pré-lecteur

Un point essentiel est que, quel que soit le niveau infra-syllabique auquel se situe réellement cet *effet phonème*, il apparaît bien dans nos résultats comme le produit d'une émergence et d'une construction. En effet, chez l'enfant de 4-5 ans, le post-effet sur l'énoncé /pe/ est significativement plus grand que le transfert aux deux voyelles de /epe/ (*effet mot*), ce qui n'est pas le cas chez les enfants plus âgés ni chez les adultes, dans cette étude. L'interprétation dans notre cadre de raisonnement est que l'enfant utiliserait plus la représentation du mot que les autres groupes d'âge dans l'articulation de la parole, et relativement moins les représentations syllabiques et infra-syllabiques. Ceci suggère que les représentations de l'enfant commencent à un niveau plutôt global (unité mot) et s'affinent progressivement vers des représentations de niveaux de granularité plus fine, dont au moins une unité infra-syllabique, qui émergerait au cours du développement.

Nous l'avons dit, plusieurs unités infra-syllabiques pourraient être invoquées. Dans ce schéma de construction progressive qui émerge de nos résultats, le geste articulatoire, au cœur de la phonologie articulatoire, n'est pas un candidat très crédible. En effet, dans la phonologie articulatoire, le geste est l'unité de base de la production. Cette unité devrait donc être présente dès l'apparition de la parole chez le bébé, avant des unités plus larges telles que le mot. On peut supposer que cette unité de base aurait au moins autant d'influence sur l'articulation chez l'enfant de 4-5 ans que chez l'adulte. Or c'est le résultat inverse qui est obtenu : *l'effet mot* est observé chez les 4-5 ans et pas chez les autres groupes. Cela tend donc à réfuter l'hypothèse

que le transfert à un niveau infra-syllabique repose sur le geste articulatoire, en tant qu'unité de base de la production.

Parmi les unités infra-syllabiques possibles, le phonème est un excellent candidat – tout comme, à ce stade, d'autres unités phonologiques comme le trait phonétique. L'apparition progressive du phonème en lien avec la dynamique des apprentissages sensorimoteurs est en cohérence avec la plupart des visions émergentistes du développement phonologique (Cristia et al., 2012; Walley et al., 2003). C'est pourquoi cette représentation infra-syllabique émergente sera par la suite désignée par le terme de représentation phonémique.

On peut se demander quand exactement apparaît cette représentation phonémique. Le transfert d'apprentissage pourrait être étudié à un âge antérieur pour répondre à cette question. Mais l'adaptation serait probablement limitante, puisqu'à 2 ans les enfants ne s'adaptent pas de manière robuste (MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012). Cette absence d'adaptation robuste pourrait être interprétée comme le fait que les unités sensorimotrices ne sont pas encore totalement constituées. Il se peut qu'une partie seulement des liens sensorimoteurs qui constituent l'unité manque à cet âge, comme la capacité à générer des prédictions sensorielles très rapides sur la base des actions effectuées (en exploitant le *feedforward model*).

En outre, nos résultats, selon lesquels l'utilisation des représentations phonémiques se renforce entre l'âge de 4-5 ans et de 7-8 ans, soulèvent une autre question : L'acquisition de la lecture joue-t-elle un rôle dans le renforcement des représentations phonémiques et leur contribution dans l'articulation de la parole ? Cette question pourrait faire l'objet de futurs travaux, qui testeraient le transfert d'apprentissage auditorimoteur au niveau phonémique chez des adultes illettrés.

5.4. Les effets lexicaux et contextuels

L'*effet mot* désigne une différence entre l'amplitude du transfert et du post-effet, révélée par l'intercept et/ou la pente des modèles rendant compte des données. Un *effet mot* a été observé chez l'adulte dans l'étude 1, ainsi que chez l'enfant, dans la 3^{ème} étude. Une analyse présentée en Annexe 1 montre que cet effet est aussi présent dans l'étude 2. Cet effet révèle une contribution du mot (et éventuellement d'unités plus larges, comme la phrase) dans le contrôle de l'articulation de la parole (voir à ce sujet Redford, 2015).

L'influence du statut lexical des stimuli d'entraînement et de test a aussi été investiguée dans l'étude 1, chez l'adulte. Nous n'avons pas observé de différence sur l'amplitude du transfert selon que le stimulus de test était un pseudo-mot ('bépé' ou 'pébé') ou bien un mot ('bébé'). L'influence du statut lexical du stimulus d'entraînement est en cours d'investigation dans une nouvelle étude que nous menons chez l'enfant. Pour cela, les amplitudes d'adaptation, de transfert et de post-effet seront contrastées entre le couple de stimuli 'fée'-'café' et le couple 'pé'-'épée'.

D'autre part, un *effet de position*, c'est-à-dire une amplitude de transfert plus élevée sur la première que sur la 2^{ème} syllabe du mot de test, a été observé dans l'étude 1. Les interprétations possibles de cet effet ont été discutées dans le chapitre 3. L'*effet de position* n'a pas été observé dans les deux autres études. Une explication est que dans ces deux études, dans lesquelles l'unique stimulus de test était 'épée' (/epe/), l'effet de position était opposé à l'*effet syllabe* attendu, qui aurait dû favoriser le transfert sur la 2^{ème} syllabe. Il est probable que ces deux effets aient eu lieu mais se soient simplement annulés mutuellement. Enfin, la modalité de présentation du mot de transfert (mot écrit ou image) a eu une influence sur l'amplitude du transfert, probablement via une influence sur l'un ou l'autre de ces deux effets. Cette influence peut être le signe d'une activation des représentations orthographiques, lors de la lecture d'un mot. Cette activation aurait une influence sur l'activation des représentations phonologiques. Cet effet de modalité peut aussi être interprété de manière plus générale comme un effet de similarité perceptuelle (*perceptual match effect*, Reder, Donavos, & Erickson, 2002), les mots écrits ayant une certaine ressemblance visuelle tandis que les images étaient totalement différentes.

Ces effets de contextes récapitulés ici (*effet mot, effet de position, effet de modalité*) révèlent des interactions entre les représentations sensorimotrices associées à l'articulation de la parole, et les niveaux lexicaux, orthographiques ou contextuels. Ces observations vont à l'encontre des propositions développées dans les modèles classiques de psycholinguistique (Dell, 1986; Levelt, 1989), plutôt issus d'une vision cognitiviste, selon laquelle la production de parole serait traitée avec des processus successifs ayant une entrée et une sortie précise, sous forme de représentations amodales, définies à un niveau linguistique donné. Nos résultats nous semblent imposer un cadre conceptuel dans lequel des liens structurants et bidirectionnels existent entre niveaux lexicaux et représentations sensorimotrices.

Des modèles inspirés des systèmes dynamiques complexes ont été développés pour rendre compte de l'influence des représentations orthographiques (Van Orden & Goldinger, 1994). La Figure 5.3 propose une illustration de deux de ces modèles. Un premier modèle rend compte de l'influence lexicale et de l'orthographe dans la reconnaissance de mots (Grainger & Ferrand, 1994; Ventura, Kolinsky, & Fernandes, 2007). Un modèle de dénomination de mots et d'images, inspiré par les résultats d'expériences d'amorçage (*priming*), a aussi été développé par le même groupe de recherche (Ferrand et al., 1994). Mais ce modèle ne prédit pas les résultats de transfert que nous avons obtenus en fonction de la modalité de présentation. En effet, il ne prévoit pas d'implication des unités sublexicales dans la dénomination d'images. Cela conduirait à l'absence de transfert aux niveaux syllabique et phonémique. Or nous avons observé un transfert significatif au niveau sublexical (de /pe/ vers /epe/) pour la dénomination d'images.

De nouveaux modèles de production de parole inspirés de ce cadre théorique pourraient être développés. Il serait alors intéressant de confronter leurs prédictions avec les résultats expérimentaux obtenus avec le paradigme d'adaptation auditorimotrice et son transfert.

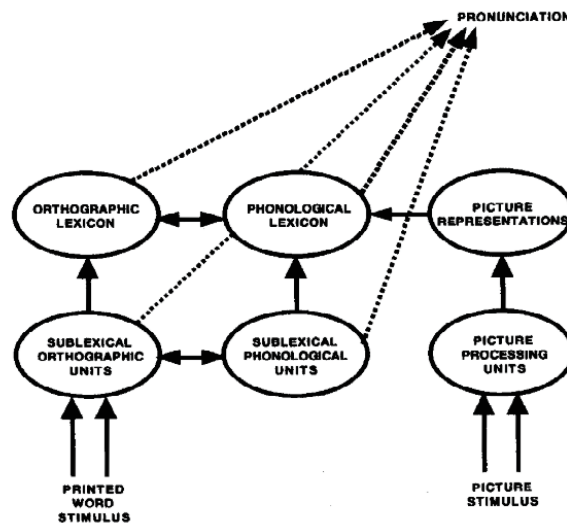


Figure 5-3: Modèle de dénomination de mots et d'images (extrait de Ferrand Grainger & Segui, 1994)

Les modèles épisodiques de la mémoire, ou modèles multi-traces, inspirés des théories de l'exemplaire, peuvent aussi fournir un cadre d'interprétation pertinent pour les phénomènes observés (Goldinger, 1998). Ces modèles reposent notamment sur l'observation d'effets de détails de surface (comme la voix qui prononce le stimulus) dans des expériences de

reconnaissance de mot (Goldinger, 1996). Dans ce cadre théorique, la perception et la mémoire reposent sur un même processus : lorsqu'un stimulus est perçu, il « résonne » avec les traces mnésiques accumulées ayant des points communs avec le stimulus perçu. Ces traces sont donc impliquées dans la perception du stimulus. Pour tester le cadre théorique de la mémoire épisodique en production de parole, le transfert d'apprentissage auditorimoteur pourrait être exploité davantage. Il serait intéressant de tester l'influence de détails de surface des stimuli, comme la police de caractère des stimuli d'entraînement et de test, ou la proximité visuelle s'il s'agit d'images à dénommer. Bien que ces détails restent associés à la perception du stimulus, leur effet sur le transfert au niveau de l'articulation de la parole permettrait de confronter ce paradigme à une vision épisodique de la parole intégrant perception et action.

5.5. Comment ces différents niveaux d'unités s'organisent-ils ?

5.5.1 Coexistence des unités

Dans notre étude chez l'adulte, nous avons mis en évidence un transfert sur 3 niveaux différents : l'énoncé ou le mot, la syllabe, et le phonème. Nous en avons déduit qu'au moins trois unités différentes contribuaient à l'articulation de la parole. Ces trois niveaux de représentations sont de nature sensorimotrice. Il n'est pas exclu que d'autres niveaux interviennent, comme la phrase, et diverses unités infra-syllabiques évoquées dans la section 5.3.

Cela va à l'encontre de la proposition développée dans le modèle de production de parole de Levelt (1989), dans lequel seuls un ou deux niveaux d'unité seraient en interface avec l'articulation de la parole. De plus, bien que certains modèles de contrôle moteur de la parole prévoient qu'il puisse exister plusieurs niveaux de correspondance entre les commandes motrices et les conséquences acoustiques attendues, ils ne précisent pas comment ces niveaux d'unités se structurent (Houde & Nagarajan, 2011; Tourville & Guenther, 2011). Ainsi cet aspect devrait être pris en compte dans les futurs travaux de modélisation, en psycholinguistique comme en contrôle moteur de la parole.

Une question majeure se pose : Quelles relations entretiennent ces différents niveaux d'unités entre eux ? Une première possibilité est que ces unités seraient organisées en une hiérarchie structurée, chaque niveau détenant un rôle bien précis. Cette possibilité est par exemple développée dans le modèle de Hickok (2012) selon lequel syllabes et phonèmes seraient

spécifiés respectivement sur les plans acoustique et somatosensoriel. Ainsi chaque niveau correspondrait à une modalité sensorielle. Néanmoins, ce modèle ne prévoit pas d'influence du mot dans le contrôle de la parole. De plus, ce type de structure ne permettrait pas d'intégrer d'autres niveaux éventuels.

Une autre possibilité est l'émergence d'une forme d'auto-organisation, où les unités de contrôle de l'articulation dépendraient entièrement de l'expérience du locuteur, et donc notamment de la fréquence d'utilisation des mots ou expressions. Par exemple, Tilsen (2016) propose que les unités de contrôle se constituent par la coordination de gestes articulatoires, en fonction de l'expérience du locuteur. Pour Redford (2015), certaines phrases fréquentes constitueraient une unité à part entière, qui n'aurait pas besoin d'être décomposée pour être articulée. A l'inverse, l'articulation de certains mots peu fréquents nécessiterait une décomposition en unités sublexicales telles que les syllabes ou les phonèmes.

Pour explorer ces hypothèses avec le paradigme d'apprentissage auditorimoteur, il faudrait probablement appliquer les perturbations et étudier le transfert non plus sur des mots isolés, mais plutôt sur des phrases entières, en contrôlant leur fréquence d'utilisation dans le langage courant. Dans une étude récente, une perturbation auditive a été appliquée en continu lors de la production d'un ensemble de phrases entières (Lametti, Smith, Watkins, & Shiller, 2018). Une adaptation a été observée sur l'ensemble des voyelles, et un transfert significatif a eu lieu vers des mots isolés. Les auteurs n'ont cependant pas cherché à distinguer les niveaux de granularité du transfert. Pour cela, il aurait peut-être fallu contraster le transfert sur des phrases de fréquences différentes, contenant ou pas les mêmes mots, ou groupes de mots. La vision de Redford (2015) prédirait un faible transfert sur les phrases très courantes, prononcées de manière globale, et un transfert plus élevé sur des phrases moins courantes, avec des mots peu fréquents. Par ailleurs, appliquer sélectivement la perturbation sur une syllabe à une position précise dans la phrase permettrait d'approfondir l'interprétation de *l'effet de position* que nous avons observé sur les items bisyllabiques.

5.5.2 Apprentissage global ou spécifique : flexibilité de l'organisation des unités

Un aspect important concernant la co-existence des unités est que leurs contributions relatives semblent être ajustables. En effet, nous avons observé que chez l'adulte, l'adaptation et le post-effet étaient significativement plus élevés que le transfert seulement lorsque l'entraînement (soumis à la perturbation auditive) était long, mais pas lorsqu'il était court (voir Annexe 1).

Cela pourrait indiquer que la contribution respective des différentes unités sensorimotrices en jeu — mots, syllabes ou phonèmes – dépend de la durée de l’entraînement. Dit autrement, l’amplitude du transfert est équivalente pour les deux durées d’entraînement, alors que l’adaptation et le post-effet augmentent avec la durée de l’entraînement. Ainsi, un entraînement court générerait un apprentissage relativement global, avec un transfert proportionnellement important, tandis qu’un entraînement plus long n’aboutirait pas à plus de transfert, mais seulement à un apprentissage spécifique à la situation d’entraînement plus important. Cette observation questionne la relation entre les différents niveaux d’unités, et suggère que leurs contributions relatives dans l’articulation de la parole sont variables.

Lametti et al. (2018) ont constaté que lorsque l’apprentissage est effectué sur des chaînes de mots dénués de sens, et non de vraies phrases, l’apprentissage était plus spécifique : un écart entre l’amplitude de l’adaptation et du transfert était observé seulement dans ce cas, alors que le transfert était d’amplitude équivalente à l’adaptation lorsque l’apprentissage se faisait avec de vraies phrases. Une situation expérimentale très stéréotypée conduirait-elle à un apprentissage plus spécifique qu’une situation plus écologique ?

D’autres travaux expérimentaux alimentent la question des relations entre les différentes unités sensorimotrices. Par exemple, des travaux sur le transfert d’apprentissage auditorimoteur ont montré qu’un même locuteur est capable d’une adaptation auditorimotrice différenciée selon le mot pour un même phonème (Rochet-Capellan & Ostry, 2011). Pour cela, une perturbation auditive différente a été appliquée aux mots anglais ‘bed’, ‘ted’ et ‘head’ pour un même locuteur, au cours d’un même entraînement. Dans une autre étude avec une seule perturbation auditive, les mêmes auteurs ont pourtant observé que l’apprentissage se transférait aux autres monosyllabes contenant le même phonème (Rochet-Capellan et al., 2012). Ces travaux suggèrent que la manipulation de l’objet de l’apprentissage peut influencer sur la spécificité ou la généralisation de l’apprentissage. Cela revient à dire que le niveau de granularité des représentations sensorimotrices modifiées par l’apprentissage auditorimoteur dépend de la situation d’apprentissage.

Des travaux similaires ont été menés chez des oiseaux, apportant un éclairage supplémentaire sur les mécanismes neuronaux permettant un apprentissage spécifique ou au contraire une généralisation plus large (Tian & Brainard, 2017). Cette étude montre que lorsque la perturbation est appliquée à une syllabe dans un contexte séquentiel donné, l’adaptation se transfère naturellement, au bout d’un certain temps, sur la même syllabe dans d’autres contextes

séquentiels. Ce transfert semble occasionné par la modification de représentations syllabiques indépendantes du contexte séquentielles, localisées dans le système moteur. A l'inverse, si une perturbation différente est appliquée dans deux contextes séquentiels différents, un signal neuronal permet de conserver la spécificité de l'apprentissage, en empêchant de modifier la représentation sensorimotrice de la syllabe.

Ainsi, nos résultats et d'autres travaux questionnent l'organisation des unités proposées dans diverses modèles de production de la parole. Ils montrent que l'implication de tel ou tel niveau d'unité dans l'apprentissage sensorimoteur dépend de l'objet de l'apprentissage et de la temporalité de celui-ci. Une meilleure compréhension de cette organisation des unités pourrait peut-être permettre d'améliorer l'efficacité de programmes de rééducation de la parole. Les effets observés suggèrent une organisation complexe des unités sensorimotrices, dont les rôles respectifs seraient ajustables en fonction du contexte. A notre connaissance, il n'existe pas à ce jour de modèle de parole permettant de rendre compte précisément des effets observés. Ce cadre reste donc à être développé, que ce soit avec une approche bayésienne, dans le prolongement du modèle COSMO (Barnaud, 2018; Moulin-Frier et al., 2015), ou bien dans le cadre des systèmes dynamiques, ou des modèles épisodiques (Goldinger & Azuma, 2003; Van Orden & Goldinger, 1994).

5.6. Conclusion générale

Cette thèse s'inscrit dans une démarche d'intégration entre la psycholinguistique et l'étude du contrôle moteur de la parole. Elle a pour objet l'étude des unités de production de la parole, aussi désignées par le terme de représentations phonologiques, et leur développement, en partant du principe que ces unités sont de nature sensorimotrice. Les unités sensorimotrices associent des informations sensorielles et motrices. Une méthode expérimentale innovante a été proposée dans ce but : un paradigme combinant l'adaptation auditorimotrice et son transfert. Le paradigme d'adaptation auditorimotrice à des perturbations de formants permet de cibler spécifiquement les unités sensorimotrices, tandis que son transfert en explore le ou les niveaux de granularité : mot, syllabe, phonème. Le transfert n'avait jamais été évalué sur des énoncés de plusieurs syllabes, ni contrasté sur différents niveaux linguistiques en production de parole.

Les résultats expérimentaux indiquent que plusieurs unités, dont le mot, la syllabe et le phonème, co-contribuent à l'articulation de la parole chez l'adulte. Ils suggèrent aussi une

influence des représentations orthographique sur ces unités. D'autre part, ils révèlent chez l'enfant pré-lecteur l'existence de représentations sensorimotrices du phonème, dont la contribution pour l'articulation de la parole se renforce entre 4-5 ans et 7-8 ans. Enfin, un lien entre adaptation auditorimotrice et conscience phonologique est mis en évidence. Ce lien valide notre hypothèse selon laquelle les représentations phonologiques, parfois qualifiées d'implicites, sont de nature sensorimotrice.

Ainsi, ces résultats remettent en cause certains aspects des modèles classiques de production de parole, et apportent de nouveaux éléments sur le développement phonologique de l'enfant. Outre leur portée sur le plan théorique, ces travaux pourraient avoir des implications cliniques, pour la dyslexie ou l'apraxie développementales, ou encore pour la rééducation de la parole suite à un accident vasculaire cérébral (AVC) par exemple. Cette thèse met en avant la pertinence de la méthode expérimentale utilisée pour explorer les représentations phonologiques qui sous-tendent la parole, et leur développement typique ou atypique.

Enfin, certaines limitations de nos travaux ouvrent de nouvelles perspectives de recherche, en particulier concernant une possible influence réciproque entre développement des représentations sensorimotrices et acquisition de la lecture. De futurs travaux sur le transfert pourraient également poursuivre les investigations concernant la flexibilité de l'organisation des unités, permettant selon les cas un apprentissage global ou spécifique.

6. Annexe : Amplitudes de l'adaptation, du transfert et du post-effet en fonction de la durée de l'entraînement

Chez l'adulte, un *effet mot*, c'est-à-dire une différence entre les amplitudes de transfert et de post-effet, a été observé dans l'étude 1 (chapitre 3), mais pas dans l'étude 3 (chapitre 4). Comment cela peut-il s'expliquer ? Nous souhaitons montrer ici que la durée de l'entraînement avec perturbation auditive a une influence sur cet effet. Pour cela, nous comparons les données des deux études du chapitre 4, réalisées avec les mêmes stimuli, /pe/ et 'épée', présentés sous forme d'image.

Un modèle linéaire mixte est bâti avec pour facteurs fixes la *phase* (adaptation, transfert, ou post-effet), la *durée* d'entraînement (long vs court), et l'*essai*. Le modèle obtenu est constitué de la somme des interactions entre phase et essai ($\text{chisq}(3)=18.2, p<0.0001$) et entre phase et durée ($\text{chisq}(3)=8.5, p=0.036$) : $\text{amplitude} \sim \text{durée} * \text{phase} + \text{essai} * \text{phase}$.

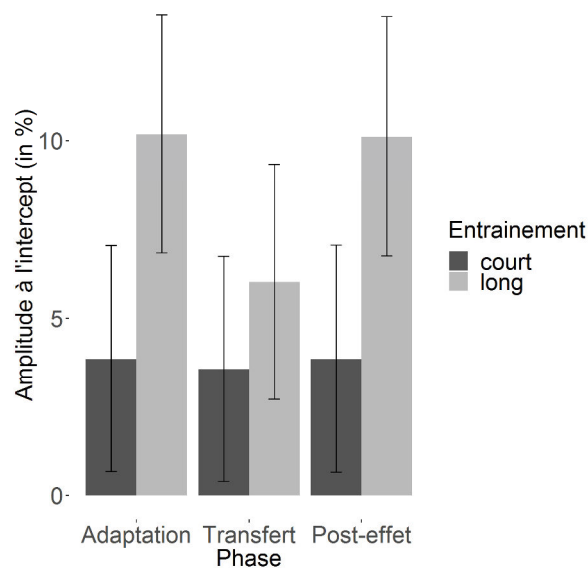


Figure 6-1: Amplitude de l'adaptation, du transfert et du post-effet, à l'intercept, selon la durée de l'entraînement avec perturbation auditive. Les barres d'erreur représentent l'intervalle de confiance à 95%.

Nous constatons, par l'analyse des intercepts, que l'adaptation et le post-effet avec un entraînement long (180 essais) sont significativement plus élevés qu'avec un entraînement court de 50 essais (respectivement $+6.2\% \pm 2.3; z=2.7; p=0.02$; $+6,3\% \pm 2.3; z=2.7; p=0.02$) tandis que les transferts sont d'amplitude équivalente ($2.4\% \pm 2.3; z=1; p=0.6$). Vu sous un autre

angle, l'adaptation et le post-effet sont significativement plus élevés que le transfert à la suite d'un entraînement long (respectivement $4.2\% \pm 1.3$; $z=3.2$; $p=0.008$; et $4\% \pm 1.3$; $z=3.2$; $p=0.01$) mais pas suite à un entraînement court ($0.3\% \pm 1.2$; $z=0.2$; $p=1$).

Par conséquent, un *effet mot* est obtenu uniquement lors d'un entraînement long. Ainsi, l'entraînement court aboutit à un apprentissage plus général, effectué sur les unités sublexicales, telles que la syllabe et le phonème. À l'inverse, un entraînement plus long n'aboutit pas à plus de généralisation, mais seulement à un apprentissage plus spécifique sur le stimulus d'entraînement.

Il est à noter qu'un autre paramètre que la durée de l'entraînement diffère entre les deux études. Il s'agit du lieu de recrutement, l'entraînement long ayant été effectué avec des français, tandis que l'entraînement court a été réalisé avec des québécois. Néanmoins, il nous semble très peu probable que l'effet observé soit lié à des différences régionales de la langue française.

7. Références

- Adrián, J. A., Alegria, J., & Morais, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology*, 30(3), 329–351. <https://doi.org/10.1080/00207599508246574>
- Ainsworth, S. (2015). *Development of phonological representations in young children*. Thèse de Doctorat, University of Manchester.
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87, 85–106. <https://doi.org/10.1016/j.jecp.2003.10.002>
- Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology*, 96(1), 43–55. <https://doi.org/10.1037/0022-0663.96.1.43>
- Araújo, J., Flanagan, S., Castro-caldas, A., & Goswami, U. (2018). The temporal modulation structure of illiterate versus literate adult speech. *PLoS ONE*, 13(10), e0205224.
- Araújo, S., Reis, A., & Petersson, K. M. (2015). Rapid Automated Naming and reading performance : A meta-analysis. *Journal of Educational Psychology*, 107(3), 868–883.
- Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society, B* 302, 311–324.
- Barbier, G. (2016). *Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice*. Thèse de Doctorat, Université Grenoble Alpes.
- Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2015). Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses. *Proceedings of Interspeech 2015*, 374–378.
- Barnaud, M.-L., Bessière, P., Diard, J., & Schwartz, J. L. (2018). Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication. *Brain and Language*, 187, 19–32. <https://doi.org/10.1016/j.bandl.2017.12.003>
- Barnaud, M. (2018). *Modélisation bayésienne du développement conjoint de la perception, l'action et la phonologie*. Thèse de Doctorat, Université Grenoble Alpes.

- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Bertelson, P., de Gelder, B., Tfouni, L. V., Morais, J., & Bertelson, P. (1989). Metaphonological abilities of adult illiterates: New evidence of heterogeneity. *European Journal of Cognitive Psychology*, 1(3), 239–250. <https://doi.org/10.1080/09541448908403083>
- Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66(4), 1001–1017. <https://doi.org/10.1121/1.383319>
- Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, 32(2), 560–570. <https://doi.org/10.1016/j.ridd.2010.12.020>
- Boets, B., Wouters, J., van Wieringen, A., De Smedt, B., & Ghesquière, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language*, 106(1), 29–40. <https://doi.org/10.1016/j.bandl.2007.12.004>
- Boets, B., Wouters, J., Van Wieringen, A., & Ghesquière, P. (2006). Auditory temporal information processing in preschool children at family risk for dyslexia: Relations with phonological abilities and developing literacy skills. *Brain and Language*, 97(1), 64–79. <https://doi.org/10.1016/j.bandl.2005.07.026>
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, 22(7), 1504–1529. <https://doi.org/10.1162/jocn>
- Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, 87, 71–83. <https://doi.org/10.1016/j.jml.2015.11.002>
- Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2–4), 140–155. <https://doi.org/10.1159/000261823>
- Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: an overview. *Phonetica*, 49, 155–180. <https://doi.org/10.1159/000261913>

- Browman, C. P., & Goldstein, L. (1995a). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion; Explorations in the dynamics of cognition* (pp. 175–193). Cambridge, MA: MIT Press.
- Browman, C. P., & Goldstein, L. (1995b). Gestural syllable position effects in American English. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues*. New York: American Institute of Physics Press.
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536. <https://doi.org/10.1073/pnas.1508631112>
- Bürki, A., Spinelli, E., & Gaskell, M. G. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, *67*(4), 449–467. <https://doi.org/10.1016/j.jml.2012.08.001>
- Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, *24*(02), 215–221. <https://doi.org/10.1017/S0272263102002061>
- Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., ... Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*(6), 678–686. <https://doi.org/10.1177/0956797611434536>
- Carroll, J. M. (2001). *The development of phonological awareness in pre-school children*. Thèse de Doctorat, University of York.
- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. [https://doi.org/10.1016/S0010-0277\(03\)00164-1](https://doi.org/10.1016/S0010-0277(03)00164-1)
- Castles, A., Coltheart, M., Wilson, K., Valpied, J., & Wedgwood, J. (2009). The genesis of reading ability: What helps children learn letter-sound correspondences? *Journal of Experimental Child Psychology*, *104*(1), 68–88. <https://doi.org/10.1016/j.jecp.2008.12.003>
- Castles, A., Wilson, K., & Coltheart, M. (2011). Early orthographic influences on phonemic awareness tasks: Evidence from a preschool training study. *Journal of Experimental Child Psychology*

Psychology, 108(1), 203–210. <https://doi.org/10.1016/j.jecp.2010.07.006>

- Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The illiterate brain: Learning to read and write during childhood influences the functional organization of the adult brain. *Brain*, 121, 1053–1063.
- Castro-Caldas, Alexandre, & Reis, A. (2003). The knowledge of orthography is a revolution in the brain. *Reading and Writing*, 16, 81–97.
- Caudrelier, T., Ménard, L., Perrier, P., Schwartz, J., Gerber, S., Vidou, C., & Rochet-Capellan, A. (2019). Transfer of sensorimotor learning reveals phoneme representations in preliterate children. *Cognition*, 192(11), 103973. <https://doi.org/10.1016/j.cognition.2019.05.010>
- Caudrelier, T., Perrier, P., Schwartz, J.-L., & Rochet-Capellan, A. (2018). Picture naming or word reading : does the modality affect speech motor adaptation and its transfer ? In *Proceedings of Interspeech 2018* (pp. 956–960).
- Caudrelier, T., & Rochet-Capellan, A. (sous presse). Changes in speech production in response to formant perturbations: An overview of two decades of research. In S. Fuchs, J. Clelland, & A. Rochet-Capellan (Eds.), *Speech production and perception: Learning and Memory*. Peter Lang.
- Caudrelier, T., Schwartz, J., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning : What does it tell us about speech production units? *Journal of Speech Language and Hearing Research*, 61(7), 1613–1625.
- Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 205–235. <https://doi.org/10.1016/j.cognition.2005.01.009>
- Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, 16(5–6), 609–636. <https://doi.org/10.1080/01690960143000083>
- Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2), 1–24. <https://doi.org/10.18637/jss.v034.i02>
- Cristia, A., & Hallé, P. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech Planning and*

Dynamics (pp. 11–38). Frankfurt am Main: Peter Lang.

- 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. <https://doi.org/10.3389/fpsyg.2012.00479>
- Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes, 12*(5/6), 839–845. <https://doi.org/10.1080/016909697386718>
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature, 304*, 159–160.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385–400.
- Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language, 46*(2), 296–322. <https://doi.org/10.1006/jmla.2001.2814>
- de Gelder, B., Vroomen, J., & Bertelson, P. (1993). The effects of alphabetic-reading competence on language representation in bilingual Chinese subjects. *Psychological Research, 55*, 315–321.
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience, 16*(4), 234–244. <https://doi.org/10.1038/nrn3924>
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321.
- DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development, 34*(4), 590–601. <https://doi.org/10.1016/j.infbeh.2011.06.005>
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149–179. <https://doi.org/10.1146/annurev.psych.55.090902.142028>
- Duncan, L. G., Colé, P., Seymour, P. H. K., & Magnan, A. (2006). Differing sequences of metaphonological development in French and English. *Journal of Child Language, 33*(2), 369–399. <https://doi.org/10.1017/S030500090600732X>
- Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by

- infants. *Perception & Psychophysics*, *16*(3), 513–521.
<https://doi.org/10.3758/BF03198580>
- Eimas, P. D., & Miller, J. L. (1980). Discrimination of information for manner of articulation. *Infant Behavior and Development*, *3*(1), 367–375. [https://doi.org/10.1016/S0163-6383\(80\)80044-0](https://doi.org/10.1016/S0163-6383(80)80044-0)
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.
- Fagan, M. K. (2009). Mean Length of Utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of Child Language*, *36*(3), 495–527. <https://doi.org/10.1017/S0305000908009070>
- Fawcett, A. J., & Nicolson, R. I. (1995). Persistent deficits in motor skill of children with dyslexia. *Journal of Motor Behavior*, *27*(3), 235–240.
<https://doi.org/10.1080/00222895.1995.9941713>
- Fawcett, A. J., & Nicolson, R. I. (2002). Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia*, *8*(4), 189–203. <https://doi.org/10.1002/dys.222>
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. <https://doi.org/10.1037/a0017196>.
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*(2), 419–439.
- Ferrand, L., Grainger, J., & Segui, J. (1994). A study of masked form priming in picture and word naming. *Memory & Cognition*, *22*(4), 431–441.
- Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, *723*(35), 708–723.
<https://doi.org/10.1006/jmla.1996.0037>
- Fodor, J. A. (1985). Précis of The Modularity of Mind. *The Behavioral and Brain Sciences*, *8*(1), 1–42.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. Retrieved from

<http://citeseer.ist.psu.edu/fodor88connectionism.html>

- Foss, D. J., & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, 12(3), 246–257. [https://doi.org/10.1016/S0022-5371\(73\)80069-6](https://doi.org/10.1016/S0022-5371(73)80069-6)
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3–28.
- Fry, D. B. (1969). The linguistic evidence of speech errors. *Brno Studies in English*, 8, 69–74.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of learning and motivation*. Vol 9. (pp. 133–177). New York: Academic Press.
- Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, 61, 191–218. <https://doi.org/10.1146/annurev.psych.093008.100408>
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3–4), 305–320. [https://doi.org/10.1016/S0095-4470\(03\)00030-5](https://doi.org/10.1016/S0095-4470(03)00030-5)
- Goswami, U. (2002). In the beginning was the rhyme? A reflection on Hulme, Hatcher, Nation, Brown, Adams, and Stuart (2002). *Journal of Experimental Child Psychology*, 82, 47–57. <https://doi.org/10.1006/jecp.2002.2673>
- Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, 16(1), 43–54. <https://doi.org/http://doi.org/10.1038/nrn3836>
- Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition

- effects of masked homophone primes. *Journal of Memory and Language*, 33, 218–233.
- Grieser, D. L., & Kuhl, P. K. (1989). Categorization of speech by infants. *Developmental Psychology*, 25(4), 577–588.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267–283. <https://doi.org/10.3758/BF03204386>
- Hämäläinen, J. A., Guttorm, T. K., Richardson, U., Alku, P., Lyytinen, H., & Leppänen, P. H. T. (2013). Auditory event-related potentials measured in kindergarten predict later reading problems at school age. *Developmental Neuropsychology*, 38(8), 550–566. <https://doi.org/10.1080/87565641.2012.718817>
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31(3–4), 373–405. <https://doi.org/10.1016/j.wocn.2003.09.006>
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28, 377–396. <https://doi.org/10.1006/jpho.2000.0121>
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145. <https://doi.org/10.1038/nrn3158>
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93(4), 411–428.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: compensation and adaptation. *Journal of Speech Language and Hearing Research*, 45(2), 295–310. <https://doi.org/1092-4388/02/4502-0295>
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5(Article 82), 1–14. <https://doi.org/10.3389/fnhum.2011.00082>
- Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, 23(6), 572–577. <https://doi.org/10.1177/0956797611435921>

- Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology*, 82(1), 2–28. <https://doi.org/10.1006/jecp.2002.2670>
- Icht, M., & Mama, Y. (2014). The production effect in memory : a prominent mnemonic in children. *Journal of Child Language*, 42(5), 1102–1124. <https://doi.org/10.1017/S0305000914000713>
- Ito, T., Coppola, J. H., & Ostry, D. J. (2016). Speech motor learning changes the neural response to both auditory and somatosensory signals. *Scientific Reports*, 6, 1–10. <https://doi.org/10.1038/srep25926>
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, 106(4), 1245–1248. <https://doi.org/10.1073/pnas.0810063106>
- Iverson, J. M., & Fagan, M. K. (2004). Infant Vocal – Motor Coordination : Precursor to the Gesture – Speech System? *Child Development*, 75(4), 1053–1066.
- Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Working Papers in Linguistics*, 50, 101–113.
- Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, 8(1), 43–52. <https://doi.org/10.1002/dys.208>
- Keren-Portnoy, T., Majorano, M., & Vihman, M. M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language*, 36(2), 235–267. <https://doi.org/10.1017/S0305000908008933>
- Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech Language and Hearing Research*, 53(5), 1280–1293.
- Klatt, D. H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279–312.
- Kolinsky, R. (1998). Spoken word recognition : A stage-processing approach to language differences. *European Journal of Cognitive Psychology*, 10(1), 1–40.
- Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate representations in spoken word

- recognition; Evidence from word illusions. *Journal of Memory and Language*, 34(1), 19–40. <https://doi.org/10.1006/jmla.1995.1002>
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107. Retrieved from <https://link-springer-com.proxy3.library.mcgill.ca/content/pdf/10.3758%2F03212211.pdf>
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831.
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of Acoustical Society of America*, 100(4 0 1), 2425–2438. https://doi.org/10.1007/978-1-4419-1698-3_1426
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), 13–21.
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44–59. <https://doi.org/10.1016/j.wocn.2010.01.002>
- Lametti, D. R., Krol, S. a, Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, 25(7), 1325–1336. <https://doi.org/10.1177/0956797614529978>
- Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust sensorimotor learning during variable sentence-level speech. *Current Biology*, 28(19), 3106–3113. <https://doi.org/10.1016/j.cub.2018.07.030>
- Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., ... Georgiou, G. K. (2018). Phonological awareness and Rapid Automated Naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading*, 00(00), 1–15. <https://doi.org/10.1080/10888438.2018.1510936>
- Lefebvre, P., & Sutton, A. (2008). Phonological awareness tasks for French-speaking preschoolers. *Canadian Journal of Speech-Language Pathology and Audiology*, 32(4),

158–168.

- Leppänen, P. H. T., Richardson, U., Eklund, K. M., & Guttorm, T. K. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Developmental Neuropsychology*, *22*(1), 407–422.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, *22*, 1–75.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. <https://doi.org/10.1037/h0044417>
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.
- Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Reducing false positives in intrusion detection by means of frequent episodes. *Perception & Psychophysics*, *62*(3), 615–625.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36. <https://doi.org/10.1097/00003446-199802000-00001>
- MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology*, *22*, 113–117. <https://doi.org/10.1016/j.cub.2011.11.052>
- MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin and Review*, *18*(6), 1197–1202. <https://doi.org/10.3758/s13423-011-0168-8>
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *The Behavioral and Brain Sciences*, *21*(4), 499–511; discussion 511–546.

<https://doi.org/10.1017/S0140525X98001265>

- MacNeilage, P. F., & Davis, B. L. (2001). The role of rhythmic cyclicities in infant action development. *Developmental Science, 1*, 79–83.
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. *Child Development, 71*(1), 153–163. <https://doi.org/10.1111/1467-8624.00129>
- MacNeilage, P. F., Davis, B. L., & Matyear, C. L. (1997). Babbling and first words : Phonetic similarities and differences, *22*, 269–277.
- Majorano, M., Vihman, M. M., & DePaolis, R. A. (2013). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204. <https://doi.org/10.1080/15475441.2013.829740>
- Malek, A., Gholizadeh, H., Amiri, S., Hekmati, I., & Pirzadeh, J. (2013). A comparative study on diadochokinetic skill of dyslexic, stuttering, and normal children. *ISRN Pediatrics, 2013*, 1–7. <https://doi.org/10.1155/2013/165193>
- Massaro, D. W. (1972). Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 79*(2), 124–145. Retrieved from <https://pdfs.semanticscholar.org/91e9/c0fa141bf585da224ac9e01dc55103aa6302.pdf>
- Mattar, A. A. G., & Ostry, D. J. (2007). Neural averaging in motor learning. *Journal of Neurophysiology, 97*(1), 220–228. <https://doi.org/10.1152/jn.00736.2006>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McGowan, R. W., McGowan, R. S., Denny, M., & Nittrouer, S. (2014). A longitudinal study of very young children's vowel production. *Journal of Speech Language and Hearing Research, 57*(1), 1–15. <https://doi.org/10.1038/nrg2484.-o>
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207–238. <https://doi.org/10.1037/0033-295X.85.3.207>
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*(2), 322–352. <https://doi.org/10.1037/a0026744>

- Ménard, L., Perrier, P., Aubin, J. M., Savariaux, C., & Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: an acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, *124*(2), 1192–1206. <https://doi.org/10.1121/1.2945704>
- Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, *25*(1), 47–56. <https://doi.org/10.1103/PhysRevD.62.024018>
- Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*, 77–92. <https://doi.org/10.1016/j.jml.2017.09.005>
- Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361. <https://doi.org/10.1016/j.cognition.2013.07.011>
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*(1–2), 45–64. [https://doi.org/10.1016/0010-0277\(86\)90004-1](https://doi.org/10.1016/0010-0277(86)90004-1)
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323–331. [https://doi.org/10.1016/0010-0277\(79\)90020-9](https://doi.org/10.1016/0010-0277(79)90020-9)
- Moulin-Frier, C., Diard, J., Schwartz, J.-L., & Bessière, P. (2015). COSMO (“Communicating about Objects using Sensory-Motor Operations”): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems.” *Journal of Phonetics*, *53*, 5–41. <https://doi.org/10.1016/j.wocn.2015.06.001>
- Mousty, P., Leybaert, J., Alegria, J., Content, A., & Morais, J. (1999). Evaluation des habiletés de lecture et d’orthographe au moyen de BELEC : Données longitudinales auprès d’enfants francophones testés en 2e et 4e années. *Revue Européenne de Psychologie Appliquée*, *49*(4), 325–342.
- Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*(5), 641–660. <https://doi.org/10.1080/01690960444000052>
- Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research?

Cognition, 35, 159–182. <https://doi.org/10.1080/00223131.2002.10875205>

- Nicolson, R. I., & Fawcett, A. J. (1995). Dyslexia is more than a phonological disability. *Dyslexia*, 1, 19–36.
- Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences*, 24(9), 508–511. [https://doi.org/10.1016/S0166-2236\(00\)01896-8](https://doi.org/10.1016/S0166-2236(00)01896-8)
- Nissen, M. J., & Bullemer, P. (1987). Attention requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1–32. [https://doi.org/10.1016/0010-0285\(87\)90002-8](https://doi.org/10.1016/0010-0285(87)90002-8)
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Mark, T. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech Language and Hearing Research*, 61(6), 1355–1368.
- Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia : A meta-analysis. *Scientific Studies of Reading*, 19(5), 340–359. <https://doi.org/10.1080/10888438.2015.1052455>
- Norris, D., & Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, 43(6), 541–550. <https://doi.org/10.3758/BF03207742>
- Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23(3), 299–370. <https://doi.org/10.1017/S0140525X00003241>
- O'Brien, G. E., McCloy, D. R., Kubota, E. C., & Yeatman, J. D. (2018). Reading ability and phoneme categorization. *Scientific Reports*, 8(1), 1–17. <https://doi.org/10.1038/s41598-018-34823-8>

- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258–278. <https://doi.org/10.1006/jmla.1993.1014>
- Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(6), 1543–1547. <https://doi.org/10.1037/a0020604>
- Patri, J.-F., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biological Cybernetics (Modeling)*, 109(6), 611–626. Retrieved from <http://hal.univ-grenoble-alpes.fr/hal-01221738>
- Patri, J.-F., Perrier, P., Schwartz, J. L., & Diard, J. (2018). What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. *PLoS Computational Biology*, 14(1), e1005942. <https://doi.org/10.1371/journal.pcbi.1005942>
- Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience*, 30(25), 8435–8444. <https://doi.org/10.1523/JNEUROSCI.5791-09.2010>
- Perre, L., Bertrand, D., Ziegler, J. C., Davis, C., & Holloway, R. (2011). Literacy affects spoken language in a non-linguistic task : an ERP study. *Frontiers in Psychology*, 2(Article 274), 1–8. <https://doi.org/10.3389/fpsyg.2011.00274>
- Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language : On-line activation or phonological restructuring ? *Brain Research*, 1275, 73–80. <https://doi.org/10.1016/j.brainres.2009.04.018>
- Perret, C., Bonin, P., & Méot, A. (2006). Syllabic priming effects in picture naming in French: Lost in the sea! *Experimental Psychology*, 53(2), 95–104.
- Perrier, P., Ma, L., & Payan, Y. (2005). Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. In *Proceedings of Interspeech 2005* (pp. 1041–1044). International Speech Communication Association, Lisbon, Portugal. Retrieved from <http://arxiv.org/abs/physics/0610170>
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In P.

- Hopper & J. Bybee (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–158). Amsterdam: John Benjamins.
- Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter. Retrieved from http://www.ling.northwestern.edu/~jbp/publications/word_specific.pdf
- Pinker, S. (1997). Words and rules in the human brain [news; comment]. *Nature*, 387(6633), 547–548.
- Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3(1), 111–132. <https://doi.org/10.1017/S0140525X00002053>
- Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences*, 27(12), 720–726.
- Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44(5), 712–722. <https://doi.org/10.1111/1469-7610.00157>
- Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, 61(1), 129–141. <https://doi.org/10.1080/17470210701508822>
- Read, C., Zhang, Y. F., Nie, H. Y., & Ding, B. Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24(1–2), 31–44. [https://doi.org/10.1016/0010-0277\(86\)90003-X](https://doi.org/10.1016/0010-0277(86)90003-X)
- Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory and Cognition*, 30(2), 312–323. <https://doi.org/10.3758/BF03195292>
- Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, 53, 141–152. <https://doi.org/10.1016/j.wocn.2015.06.006>
- Reis, A., & Castro-Caldas, A. (1997). Illiteracy: a cause for biased cognitive development. *Journal of the International Neuropsychological Society : JINS*, 3(5), 444–450. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9322403>

- Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomical Embryology*, 210, 419–421. <https://doi.org/10.1007/s00429-005-0039-z>
- Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: New concepts. *Electroencephalography and Clinical Neurophysiology*, 106(4), 283–296. [https://doi.org/10.1016/S0013-4694\(98\)00022-4](https://doi.org/10.1016/S0013-4694(98)00022-4)
- Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory-motor transformations in speech. *The Journal of Neuroscience*, 31(7), 2657–2662. <https://doi.org/10.1523/JNEUROSCI.6020-10.2011>
- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717. <https://doi.org/10.1152/jn.00773.2011>
- Romeo, R., & Hazan, V. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of Acoustical Society of America*, 134(5), 3781–3792. <https://doi.org/10.1121/1.4824160>
- Rousset, I. (2008). *Structures syllabiques et lexicales des langues du monde: Données, typologie, tendances universelles et contraintes substantielles*. Thèse de Doctorat, Université de Grenoble.
- Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19(5), 394–398. <https://doi.org/10.3758/BF03199398>
- Saletta, M. (2015). Literacy transforms speech production. *Frontiers in Psychology*, 6, 1458. <https://doi.org/10.3389/fpsyg.2015.01458>
- Saletta, M., Goffman, L., & Brentari, D. (2015). Reading skill and exposure to orthography influence speech production. *Applied Psycholinguistics*, 1–24.
- Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, 98(5), 2428–2442. <https://doi.org/10.1121/1.413277>
- Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of*

Verbal Learning and Verbal Behavior Learning, 9(3), 295–302.

- Scheerer, N. E., & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience Letters*, 529, 128–132. <https://doi.org/10.1016/j.neulet.2012.09.012>
- Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336–354. <https://doi.org/10.1016/j.jneuroling.2009.12.004>
- Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2), B69–B78. <https://doi.org/10.1016/j.cognition.2004.01.006>
- Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, 31(1), 18–32. [https://doi.org/10.1016/0749-596X\(92\)90003-G](https://doi.org/10.1016/0749-596X(92)90003-G)
- Segui, J., Frauenfelder, U. H., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, 72(4), 471–477. <https://doi.org/DOI:10.1111/j.2044-8295.1981.tb01776.x>
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 546–554.
- Serniclaes, W., Heghe, S. Van, Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, 87, 336–361.
- Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, 98(2), 35–44. <https://doi.org/10.1016/j.cognition.2005.03.002>
- Shadmehr, R. (2012). Generalization as a behavioral window to the neural mechanisms of learning internal models. *Human Mov Science*, 29(6), 997–1003. <https://doi.org/10.1016/j.humov.2004.04.003>
- Shapiro, L. R., Carroll, J. M., & Solity, J. E. (2013). Separating the influences of prereading skills on early word and nonword reading. *Journal of Experimental Child Psychology*,

116(2), 278–295. <https://doi.org/10.1016/j.jecp.2013.05.011>

- Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 41–55. [https://doi.org/10.1016/S0022-5371\(79\)90554-1](https://doi.org/10.1016/S0022-5371(79)90554-1)
- Shiller, D. M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4). <https://doi.org/10.1037/a0036660>
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105. <https://doi.org/10.1016/j.bandl.2016.10.004>
- Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399. <https://doi.org/10.1093/cercor/bhl147>
- Smith, B. L., Brown-Sweeney, S., & Stoel-Gammon, C. (1989). A quantitative analysis of reduplicated and variegated babbling. *First Language*, 9(6), 175–189. <https://doi.org/10.1177/014272378900900605>
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343–348. [https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6)
- Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech Language and Hearing Research*, 55(2), 596–608. [https://doi.org/10.1044/1092-4388\(2011/10-0113\).Lexical](https://doi.org/10.1044/1092-4388(2011/10-0113).Lexical)
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48(2), 233–254. [https://doi.org/10.1016/S0749-596X\(02\)00513-2](https://doi.org/10.1016/S0749-596X(02)00513-2)
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception

- than in word-learning tasks. *Nature*, 388, 381–382.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358–1368. <https://doi.org/10.1121/1.382102>
- Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38(01), 1–34.
- Stoodley, C. J., Harrison, E. P. D., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, 44(5), 795–798. <https://doi.org/10.1016/j.neuropsychologia.2005.07.009>
- Swingle, D. (2003). Phonetic detail in the developing lexicon. *Journal of Speech, Language & Hearing Research*, 46(2–3), 265–294.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9(2), 182–198. [https://doi.org/10.1016/0093-934X\(80\)90139-X](https://doi.org/10.1016/0093-934X(80)90139-X)
- Terband, H., van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, 51, 64–77. <https://doi.org/10.1016/j.jcomdis.2014.06.009>
- Thelen, E. (1981). Kicking, rocking, and waving: Contextual analysis of rhythmical stereotypies in normal human infants. *Animal Behaviour*, 29(1), 3–11. [https://doi.org/10.1016/S0003-3472\(81\)80146-7](https://doi.org/10.1016/S0003-3472(81)80146-7)
- Thelen, E., Schönner, G., & Scheier, C. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*.
- Tian, L. Y., & Brainard, M. S. (2017). Discrete circuits support generalized versus context-specific vocal learning in the songbird. *Neuron*, 96(5), 1168–1177. <https://doi.org/10.1016/j.neuron.2017.10.019>
- Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech

- planning. *PLoS ONE*, 8(4), e62800. <https://doi.org/10.1371/journal.pone.0062800>
- Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55, 53–77. <https://doi.org/10.1016/j.wocn.2015.11.005>
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>.
- Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9(1), 91–96. <https://doi.org/10.1037/h0034999>
- Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47(2), 466–472. <https://doi.org/10.2307/1128803>
- Trehub, S. E., & Rabinovitch, M. S. (1972). Auditory-linguistic sensitivity in early infancy. *Developmental Psychology*, 6(1), 74–77. <https://doi.org/10.1037/h0032197>
- Treiman, R., & Baron, J. (1983). Phonemic-analysis training helps children benefit from spelling-sound rules. *Memory & Cognition*, 11(4), 382–389. <https://doi.org/10.3758/BF03202453>
- Treiman, R., & Cassar, M. (1997). Can children and adults focus on sound as opposed to spelling in a phoneme counting task? *Developmental Psychology*, 33(5), 771–780. <https://doi.org/10.1037/0012-1649.33.5.771>
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866–869. <https://doi.org/10.1038/nature01710>
- Välilmaa-Blum, R. (2009). The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, 2. <https://doi.org/10.4000/cognitextes.211>
- Van Den Bunt, M. R., Groen, M. A., Frost, S., Lau, A., Preston, J. L., Gracco, V. L., ... Verhoeven, L. T. W. (2018). Sensorimotor control of speech and children's reading ability. *Scientific Studies of Reading*, 22(6), 1–14. <https://doi.org/10.1080/10888438.2018.1491583>
- Van Den Bunt, M. R., Groen, M. A., Ito, T., Francisco, A. A., Gracco, V. L., Pugh, K. R., &

- Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit. *Journal of Speech Language and Hearing Research*, *60*(3), 654. https://doi.org/10.1044/2016_JSLHR-L-16-0201
- Van Den Bunt, M. R., Groen, M. A., van der Kleij, S. W., Noordenbos, M. W., Segers, E., Pugh, K. R., & Verhoeven, L. (2018). Deficient response to altered auditory feedback in dyslexia. *Developmental Neuropsychology*, *43*(7), 622–641. <https://doi.org/10.1080/87565641.2018.1495723>
- Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(6), 1269–1291.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge: MIT Press.
- Ventura, P., Kolinsky, R., Brito-Mendes, C., & Morais, J. (2001). Mental representations of the syllable internal structure are influenced by orthography. *Language and Cognitive Processes*, *16*(4), 393–418. <https://doi.org/10.1080/01690960042000184>
- Ventura, P., Kolinsky, R., & Fernandes, J. Q. S. (2007). Is phonological encoding in naming influenced by literacy? *Journal of Psycholinguistic Research*, *36*(5), 341–360. <https://doi.org/10.1007/s10936-006-9048-1>
- Ventura, P., Kolinsky, R., Fernandes, S., Querido, L., & Morais, J. (2007). Lexical restructuring in the absence of literacy. *Cognition*, *105*(2), 334–361. <https://doi.org/10.1016/j.cognition.2006.10.002>
- Versace, R., Labeye, É., Badard, G., & Rose, M. (2009). The contents of long-term memory and the emergence of knowledge. *European Journal of Cognitive Psychology*, *21*(4), 522–560. <https://doi.org/10.1080/09541440801951844>
- Viana, A. R., Razuk, M., de Freitas, P. B., & Barela, J. A. (2013). Sensorimotor integration in dyslexic children under different sensory stimulations. *PLoS ONE*, *8*(8), e72719. <https://doi.org/10.1371/journal.pone.0072719>
- Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, *45*(4), 683–725. <https://doi.org/10.1515/LING.2007.021>

- Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2009). A dynamic systems approach to babbling and words. In E. L. Bavin (Ed.), *The Cambridge Handbook of Child Language* (pp. 163–184). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511576164.010>
- Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*(s2), 121–140. <https://doi.org/10.1111/lang.12058>
- Vihman, M. M., & McCune, L. (1994). When is a word a word? *Journal of Child Language*, *21*, 517–542.
- Vilain, A., Dole, M., Loevenbruck, H., Pascalis, O., & Schwartz, J.-L. (n.d.). The role of production abilities in the perception of consonant category in infants. *Developmental Science*.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., ... Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479. <https://doi.org/10.1037/0012-1649.33.3.468>
- Walley, A. C., & Flege, J. E. (1999). Effects of lexical status on native and non-native vowel perception: A developmental study. *Journal of Phonetics*, *27*(3), 307–332.
- Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, *16*(1), 5–20. <https://doi.org/10.1023/A:1021789804977>
- Wang, J., Wong, A. W. K., Wang, S., & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports*, *7*(1), 1–8. <https://doi.org/10.1038/s41598-017-06186-z>
- Wauquier, S. (2006). Du son au sens: acquérir ou apprendre la phonologie? *Recherches Linguistiques de Vincennes*, *35*, 5–30. <https://doi.org/10.4000/rlv.1411>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.
- White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., & Ramus, F. (2006). The

- role of sensorimotor impairment in dyslexia: A multiple case study of dyslexic children. *Developmental Science*, 9(3), 237–255. <https://doi.org/10.1111/j.1467-7687.2006.00483.x>
- Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9), 338–347. [https://doi.org/10.1016/S1364-6613\(98\)01221-2](https://doi.org/10.1016/S1364-6613(98)01221-2)
- Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3), 213–232. <https://doi.org/10.1037/h0044155>
- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, 24(5), 603–612. <https://doi.org/10.1177/0956797612458802>
- Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42(2), 193–208. <https://doi.org/10.1017/S0025100312000060>
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 48(4), 779–793. [https://doi.org/10.1016/S0749-596X\(03\)00006-8](https://doi.org/10.1016/S0749-596X(03)00006-8)
- Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New insights on developmental dyslexia subtypes: Heterogeneity of mixed reading profiles. *PLoS ONE*, 9(6), e99337. <https://doi.org/10.1371/journal.pone.0099337>
- Zoubrinetzky, R., Collet, G., Serniclaes, W., Nguyen-Morel, M. A., & Valdois, S. (2016). Relationships between categorical perception of phonemes, phoneme awareness, and visual attention span in developmental dyslexia. *PLoS ONE*, 11(3), e0151015. <https://doi.org/10.1371/journal.pone.0151015>
- Zwitserslood, P., Schriefers, H., Lahiri, A., & van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 260–271. <https://doi.org/10.1037/0278-7393.19.2.260>