



UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE 356 COGNITION LANGAGE EDUCATION

INSTITUTE OF LANGUAGE COMMUNICATION AND THE BRAIN

LABORATOIRE PAROLE ET LANGAGE/UMR 7309 CNRS AMU

INSTITUT DE NEUROSCIENCES DES SYSTEMES/INSERM

Thèse présentée pour obtenir le grade universitaire de docteur en

Sciences du Langage

Céline HIDALGO

Rythme de parole dans l'interaction langagière : bénéfice d'un entraînement rythmique musical chez l'enfant sourd

Soutenue le 20/12/2018 devant le jury :

Dr. Barbara TILLMANN	CRNL, Lyon	Rapporteur
Pr. Maya GRATIER	Université Paris Nanterre	Rapporteur
Dr. Jean-Luc SCHWARTZ	GIPSA-lab, Grenoble	Examineur
Pr. Noël NGUYEN	LPL, Aix-en-Provence	Directeur de thèse
Dr. Daniele SCHÖN	INS, Marseille	Co-directeur de thèse

Résumé

La musique et la parole requièrent l'élaboration d'informations acoustiques qui changent dans le temps. De plus, la parole, et bien plus encore la musique, possèdent toutes deux un certain degré d'organisation temporelle, un certain degré de régularité dans le temps qui leur confère un caractère rythmique. Les stimuli de nature rythmique ont la particularité de pouvoir être anticipés par le cerveau et des études en linguistique et neurosciences ont montré que plus le cerveau est capable d'anticiper les événements auditifs, meilleure est la qualité du traitement des stimuli. Les enfants sourds, bien que bénéficiant d'un input auditif de plus en plus précis grâce aux implants cochléaires ou aux prothèses conventionnelles et d'une prise en charge précoce, n'atteignent pas des niveaux de langage homogènes et souffrent toujours de difficultés de perception en milieux bruyants ou lors de conversations quotidiennes à plusieurs. La situation conversationnelle présente un contexte complexe, nécessitant l'activation de la voie audio-motrice pour anticiper et s'adapter aux variations de la parole de son interlocuteur notamment au niveau temporel. Dans ce travail de thèse, nous avons cherché à analyser, grâce à des mesures électrophysiologiques et comportementales, si un entraînement rythmique actif d'une durée de 30 minutes, stimulant la capacité d'anticipation de structures temporelles, pouvait avoir un effet sur les capacités de perception et d'accommodation temporelles de l'enfant sourd dans une tâche de dénomination en alternance avec un partenaire virtuel. Nous avons également testé les capacités rythmiques de ces enfants à différents niveaux de complexités acoustiques et structurales afin de mieux comprendre leurs déficits sous-jacents de perception temporelle de la parole. Les résultats montrent que les enfants sourds souffrent de difficultés à structurer les événements acoustiques selon différents niveaux de hiérarchie (Hidalgo, Truy & Schön, en préparation) mais qu'un entraînement rythmique de 30 minutes versus une stimulation linguistique ou auditive, permet d'améliorer leurs compétences de perception et de production temporelles de la parole dans une situation d'interaction (Hidalgo, Falk & Schön, 2017; Hidalgo, Pesnot -- Lerousseau, Marquis, Roman, Nguyen, Schön, soumis).

Abstract

Music and speech are acoustic information that evolve through time. Moreover, speech, and still more music, both possess a certain degree of temporal organization they share the property of being rhythmic, i.e. they have a certain degree of regularity across time. Numerous studies in linguistics and neuroscience have shown that the brain is able to extract these regularities and use them to anticipate the forthcoming stimuli. It is furthermore well established that the better the brain is able to anticipate auditory events, the better the quality of stimulus processing. Deaf children benefit from more and more precise auditory inputs due to great advances in cochlear implants and hearing aids development, together with early rehabilitation interventions. However, a great majority of them do not achieve consistent language levels and still have strong difficulties in noisy environments or daily conversations. The conversational situation presents a complex context, requiring the activation of the audio-motor path to anticipate and adapt to the variations of the speech of its interlocutor notably at the temporal level. In this thesis work, we have first investigated the temporal perception and accommodation capacities of deaf children in a naming task alternating with a virtual partner, at both behavioral and electrophysiological levels. We have also tested whether an active rhythmic training lasting 30 minutes, stimulating the temporal anticipation capacity, could enhance these conversational abilities. Then, we have investigated the rhythmic abilities of these children at different levels of acoustic and structural complexities to better understand their deficits. The results show that deaf children suffer from difficulties in structuring acoustic events according to different levels of hierarchy (Hidalgo, Truy & Schön, in preparation) but that a rhythmic training of 30 minutes versus a linguistic or auditory stimulation, makes it possible to improve their skills of temporal perception and production of speech in a situation of interaction (Hidalgo, Falk & Schön, 2017, Hidalgo, Pesnot--Lerousseau, Marquis, Roman, Nguyen, Schön, submitted).

Table des matières

Partie Théorique	1
I. Le rythme, une propriété importante pour la perception de la parole.....	3
1. Rythme de parole : un rythme musical ?.....	3
2. <i>L'entraînement</i> rythmique : ou comment utiliser les propriétés de la temporalité de la parole et du fonctionnement neuronal pour améliorer la perception ?.....	10
3. La pratique musicale : une activité rythmique qui améliore la perception de la parole.....	18
II. L'interaction sociale : un contexte ou le rythme (de parole) est essentiel	20
1. Le tour de parole, un contexte interactionnel qui nécessite prédiction temporelle et coordination.....	20
a. La composante temporelle du tour de parole.....	21
b. Les prédictions temporelles dans le tour de parole.....	23
2. Les mécanismes d'anticipation et d'adaptation dans l'interaction sociale	27
a. La convergence	27
b. Le phénomène d' <i>entraînement</i> sur la parole.....	28
c. La voie sensori-motrice et les émulateurs internes.....	30
d. Synchronisation brain to brain.....	37
3. Un modèle d'explication de l'adaptation dans l'interaction : ADAM.....	40
4. Les effets de la pratique du rythme musical sur les prédictions et l'adaptation temporelles en situation d'interaction.....	41
III. Le développement normal de la parole, un processus qui passe par le rythme de l'interaction	44
1. Développement précoce de la perception et de la production de structures temporelles..	44
2. De l'émergence de la rythmicité et des prédictions dans l'interaction	49

IV. Les capacités de prédictions des enfants sourds, une pièce manquante dans l'explication des déficits perceptifs des enfants sourds.....	54
1. Mise en place, fonctionnement et perception auditive avec un l'implant cochléaire.....	54
2. Conséquences de la surdit� et de l'implantation cochl�aire sur le d�veloppement du langage.	59
3. Cons�quences de la surdit� profonde cong�nitale sur le d�veloppement de l'organisation et du fonctionnement des structures c�r�brales et des processus de traitement de l'information sensorielle	64
4. Effet de la stimulation musicale sur la perception auditive et la production de la parole chez l'enfant sourd.....	71
 Partie Exp�rimentale.....	77
Exp�rience n�1	80
Exp�rience n�2	97
Exp�rience n�3	106
 Partie Discussion.....	150
V. Rappel des r�sultats.....	151
1. Y-a-t-il une probl�matique de traitement temporel chez l'enfant sourd et si oui, comment la caract�riser ?	156
a. Pr�servation de la capacit� de discrimination/reproduction relative de dur�es dans la parole.....	156
b. Perception temporelle moins pr�cise de la r�gularit� de la pulsation.....	157
c. Difficult�s de perception d'une structure temporelle de nature hi�rarchique.....	160
2. Quel type d'am�lioration une stimulation rythmique de 30 minutes apporte-t-elle aux enfants sourds en termes de traitement temporel de la parole en situation d'interaction ? Un point de vue selon le mod�le ADAM.	169
3. Est-ce que les th�ories pr�nant le rythme/musique comme outil de stimulation pour le langage sont pertinentes dans le cadre de la surdit� ?	175
4. Limites et Perspectives.....	181
 R�f�rences	185

Partie Théorique

I. Le rythme, une propriété importante pour la perception de la parole

1. Rythme de parole : un rythme musical ?

D'un point de vue physique, la musique comme le langage sont composés de sons complexes de différentes fréquences, intensités et durées qui se succèdent dans le temps. Autrement dit, lorsque nous écoutons de la musique ou la parole d'un interlocuteur, notre cerveau reçoit, en continu, un flux de stimulations acoustiques. Pour que ces deux activités fassent sens, notre cerveau doit en première instance parvenir à segmenter temporellement ce flux quasi-continu en unités qui seront pertinentes pour l'appréciation de la musique et la compréhension du langage. Sans le rythme, un auditeur ne pourrait pas créer du sens à partir des notes de musique qui se succèdent les unes après les autres, ni comprendre aisément le sens d'un discours.

Afin de structurer l'information auditive lors de l'écoute de la musique, le cerveau aura tendance à percevoir les durées des et surtout les intervalles temporels qui les séparent sons comme réguliers. Ainsi, les éléments temporels, i.e. notes ou intervalles, ne vont pas être appréhendés en fonction de leur durée absolue mais selon les rapports de régularité qu'ils entretiennent. Par exemple, les notes de musique ne vont pas être classées selon leur durée réelle en millisecondes (e.g. Do = 900 millisecondes et Si = 450 millisecondes) mais seront catégorisées selon les rapports simples de durées. En reprenant l'exemple ci-dessus, le Do sera considéré comme étant deux fois plus long que le Si. Le Do pourrait aussi durer un peu plus ou bien un peu moins et être encore perçu comme le double de la durée du Si. C'est grâce à ces rapports de durées des événements et/ou entre les événements que nous pouvons reconnaître des rythmes familiers lorsqu'ils sont joués à des tempi différents.

Dans la parole, la sensation de rythmicité est également présente mais cette rythmicité sera différemment perçue en fonction des langues. Des langues comme l'Anglais, le Russe ou l'Allemand ont été opposées, en termes de rythme, à des langues comme l'Espagnol ou le Français. Pike (1947) et Abercrombie (1967) ont ainsi défini deux groupes rythmiques de langues : des langues dont le rythme serait basé sur l'apparition régulière

d'accents, dites *stress-timed*, et d'autres langues basées sur l'apparition périodique des syllabes dites *syllable-timed*. Une troisième classe de langues, nommée *mora-timed* et marquée par des intervalles réguliers entre des consonnes géminées ou des voyelles longues (i.e. Honda contient deux syllabes mais trois mora : ho-n-da), a également été répertoriée ; cette classe contenait des langues comme le japonais (Pike, 1947; Abercrombie, 1967). Cependant, les études acoustico-phonétiques ne sont pas parvenues à démontrer une stricte régularité d'apparition autrement dit *l'isochronie* des accent-syllabes-mora dans les différentes classes de langues comme pour la musique (Dauer, 1983). En se basant sur la notion de durées relatives, de nouvelles recherches ont tenté d'analyser le concept de régularité mais à partir des durées des voyelles et des consonnes (Ramus, Nespors, & Mehler, 1999) ainsi que la durée des intervalles qui sépare le début d'émission de chaque phonème dans différentes langues (Grabe & Low, 2002). Cette mesure, nommée Pairwise Variability Index, a permis de réaliser un nouveau classement des langues et bien que certaines langues semblent appartenir à un groupe commun, la catégorisation des langues en classes rythmiques distinctes selon ces critères, reste à appréhender avec précaution (Cummins, 2002). Cependant, la nature perceptuelle du rythme a poussé les chercheurs à mesurer la régularité non pas directement dans le signal acoustique mais plutôt dans le comportement des auditeurs, pour tenter de définir l'indice qui permettrait de percevoir une rythmicité dans la parole. Ainsi, des études réalisées dans les années 70-80 (Marcus, 1981; Morton, Marcus, & Frankish, 1976) ont montré que lorsque des auditeurs doivent déterminer si des séquences de chiffres sont prononcées régulièrement ou pas ou produire des séquences de syllabes en rythme sur un métronome, ils n'utilisent pas le début du mot, i.e. *l'onset*, pour déterminer une régularité des intervalles dans la parole mais un point situé approximativement dans les 2/3 de la montée d'amplitude de la voyelle de chaque syllabe. C'est à partir des intervalles qui séparent ces points "perceptuels" nommés *perceptual-centers* ou *p-centers*, que les auditeurs parviendraient à percevoir et produire une régularité dans la parole, donc à ressentir ce qu'on peut assimiler à une pulsation dans la musique.

La pulsation ou *beat* en musique est une structure temporelle qui émerge perceptuellement de l'organisation régulière des durées entre les notes autrement

dit du rythme de surface. En régularisant les rapports de durée entre les notes, le système perceptif génère une structure régulière de base, une période de référence qui est maintenue même en l'absence d'évènement dans le signal acoustique (figure 1.1).

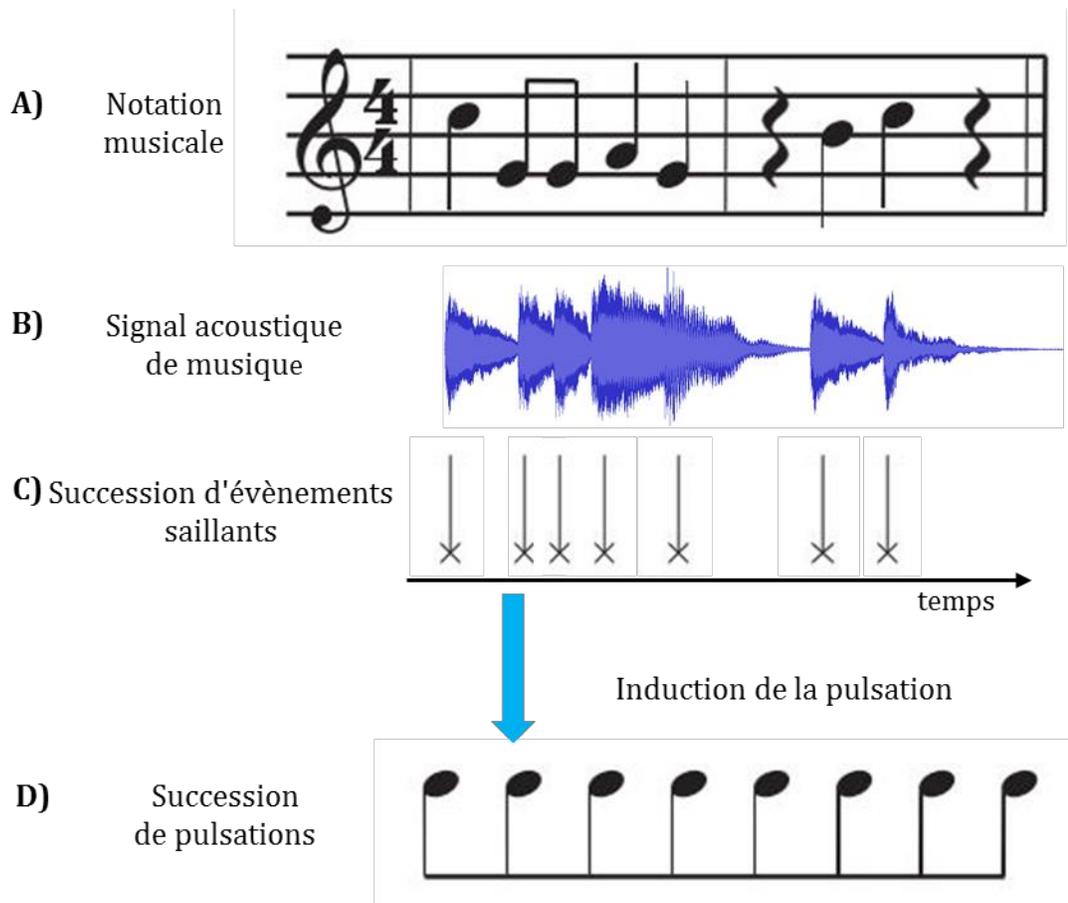


Figure 1.1. Représentation de l'induction de la pulsation adaptée de Fitch (2013). Panel A) Notation du rythme musical dans lequel sont insérées des silences. Panel B) Signal acoustique musical duquel ressort une succession d'évènement saillants (Panel C). Panel D) depuis ces évènements saillants, notre système perceptif extrait une régularité d'apparition des évènements, la *pulsation*, qui pourra être perçue même en l'absence d'une matérialisation dans le signal acoustique (voir les silences dans la notation musicale) (Fitch, 2013).

Selon la Dynamic Attending Theory (DAT) (Large & Jones, 1999), l'émergence de la pulsation proviendrait de la capacité du système perceptif à distribuer l'énergie de l'attention sur des points perçus comme saillants dans le rythme musical grâce à un

mécanisme de couplage de nature oscillatoire entre le cerveau et le signal acoustique (Figure 1.2). Une fois l'oscillateur en phase avec le signal musical, celui-ci continuerait à osciller régulièrement ce qui maintiendrait l'orientation de l'attention dans le temps même en l'absence de signal.

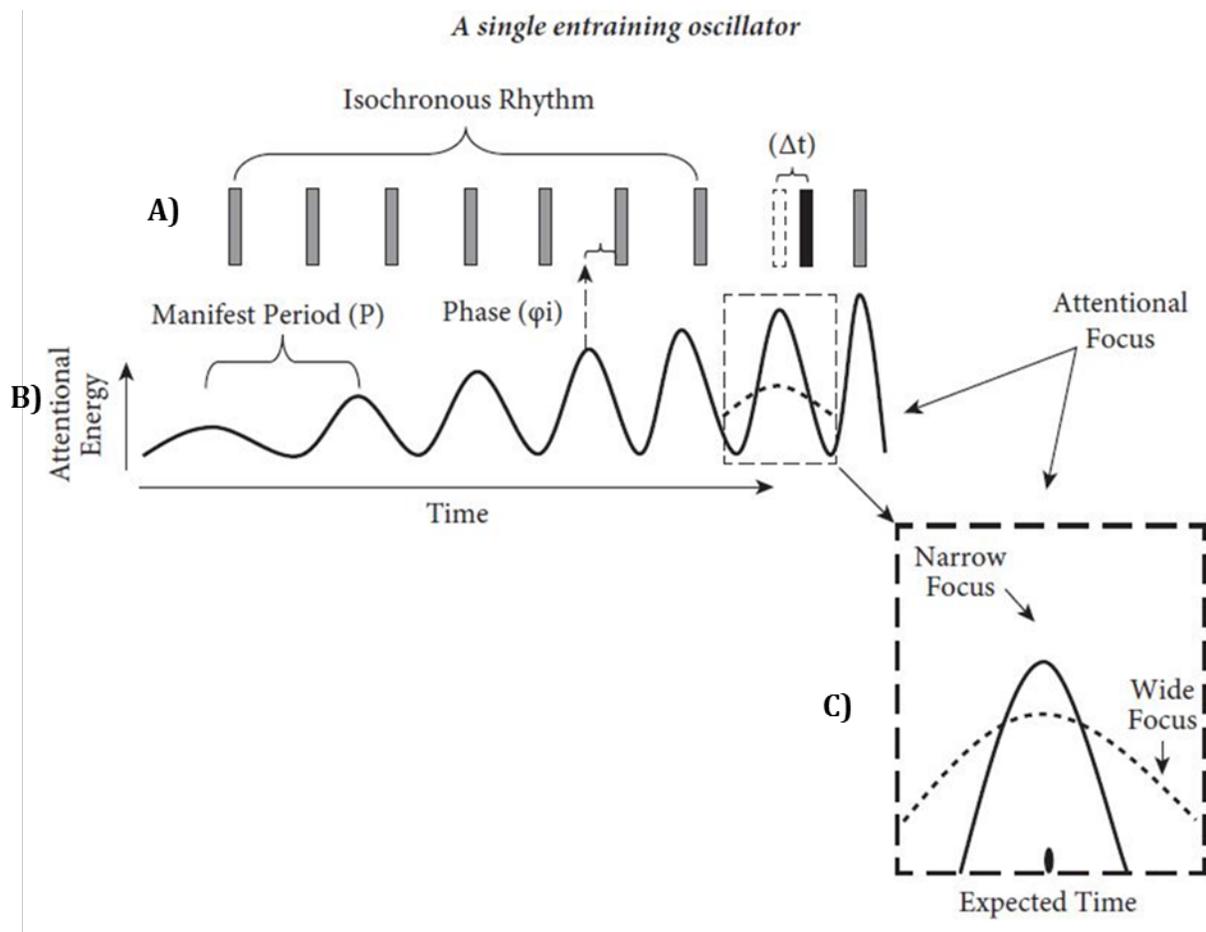


Figure 1.2. Représentation de la synchronisation entre un oscillateur et un stimulus acoustique isochrone extraite de Jones (2016). Panel A) Stimulus isochrone. Panel B) Oscillateur dont l'amplitude de la phase détermine la concentration d'énergie attentionnelle. La phase de l'oscillateur va se recalculer en fonction de l'attente du prochain événement du stimulus. Panel C), lorsque l'oscillateur a suffisamment ajusté sa phase avec le stimulus, la concentration de l'énergie attentionnelle est moins dispersée (*wide focus*) et devient maximale sur les événements du stimulus (*narrow focus*). L'oscillateur en oscillant sur ce

nouveau mode synchronisé au stimulus, peut anticiper le prochain évènement en extrapolant son futur moment d'occurrence (Jones, 2016).

L'Attentional Bounce Hypothesis (ABH) (Pitt & Samuel, 1990) considère que le même type de mécanisme est actif lors de la perception de la parole: notre attention se porterait sur des évènements perceptivement saillants et réguliers du signal, en l'occurrence les p-centers, ce qui nous donnerait, selon le même type de mécanisme oscillatoire, la sensation d'une pulsation dans la parole mais plus spécifiquement sur les p-centers contenus dans les syllabes accentuées.

La sensation de rythmicité serait en effet également affectée par une autre structure temporelle de nature perceptive, la métrique. En musique, la métrique est un niveau de structuration du rythme, plus complexe que la perception de la pulsation. La perception de la métrique se situerait à un niveau plus profond de structuration temporelle du signal et consisterait à assembler les pulsations selon une alternance d'accents faibles et d'accents forts d'où émergeraient des patterns d'accents de différentes durées, intégrées les unes dans les autres de manière hiérarchique. La structuration hiérarchique de l'information acoustique en patterns accentuels ferait ressentir certaines pulsations, celles situées dans les hauts niveaux de la hiérarchie, comme plus saillantes que d'autres (figure 1.3).

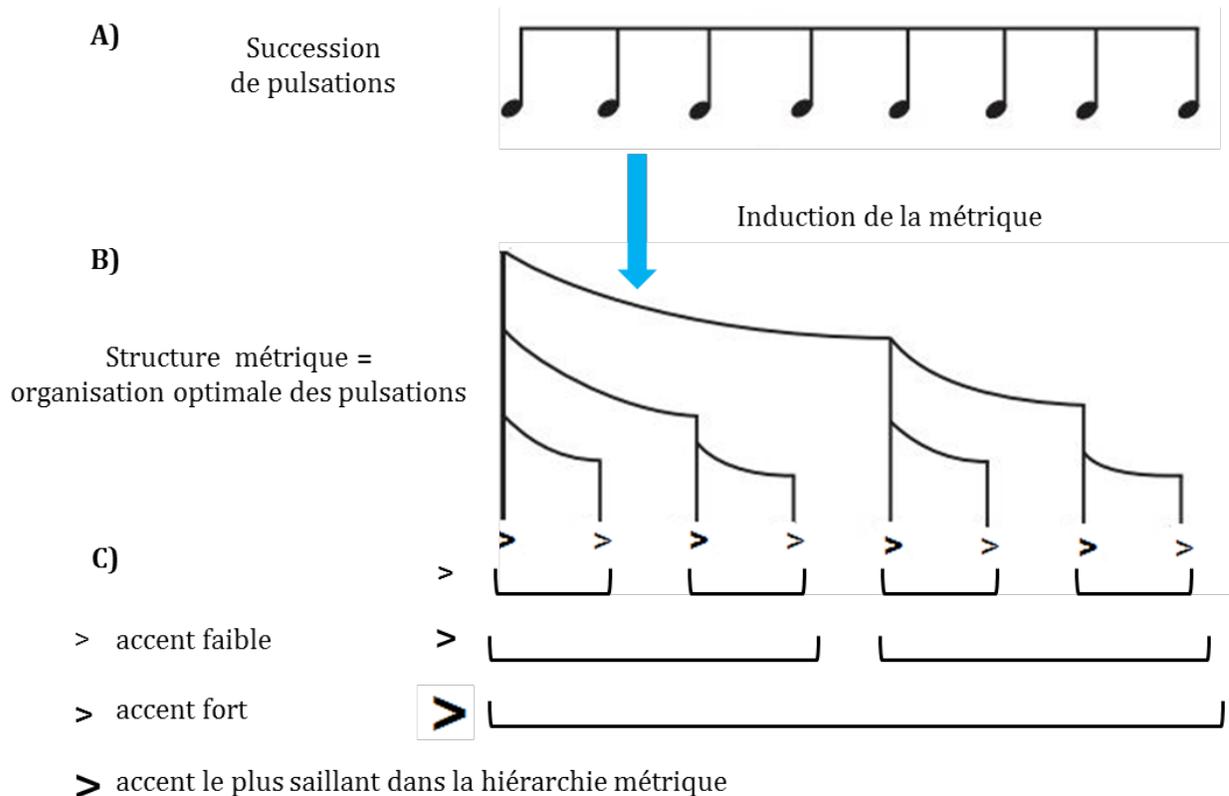


Figure 1.3. Représentation de l'émergence de la métrique. Panel A) les pulsations successives sont groupées en patterns d'accents fort/faible à différents niveaux (panel B). Panel C) ces patterns d'accents sont ensuite structurés selon différents niveaux de hiérarchie. L'auditeur perçoit ainsi une structuration optimale de l'input acoustique à différentes échelles temporelles.

Dans la parole, il est également possible de considérer une structure métrique qui émanerait de l'organisation de syllabes accentuées (notée X) versus non accentuées notées (x). On sait par exemple qu'il y a des langues donc l'accent des mots se porte généralement à gauche ce qui donne un pattern métrique de type iambique Xx et d'autres langues, qui ont plutôt un pattern métrique de type trochaïque xX avec l'accent à droite (Vaissière, 1991). Ce type de structure récurrente dans une langue permet de fixer un autre niveau d'organisation des p-centers. En outre, si l'on considère l'organisation des constituants prosodiques du français selon la classification de Jun et Fougeron (2002), on peut remarquer qu'ils sont

également organisés de manière hiérarchique selon différents niveaux allant de la syllabe au syntagme intonatif avec, pour chacun des niveaux, un degré d'accentuation correspondant (voir figure 1.4) (Jun & Fougeron, 2002). Dans le domaine de la prosodie, l'accentuation en français est dite démarcative et a pour fonction de marquer la frontière droite de chaque constituant de l'énoncé afin de grouper les éléments en unités de sens. L'accent est réalisé par un allongement de la durée de la syllabe et cet allongement sera plus ou moins important en fonction du niveau hiérarchique du constituant dans l'énoncé ; la dernière syllabe du dernier constituant de l'énoncé dans la hiérarchie, le syntagme intonatif, sera par exemple celle qui recevra l'accent le plus important. Cette organisation permet de constituer une grille métrique des énoncés (voir figure 1.4).

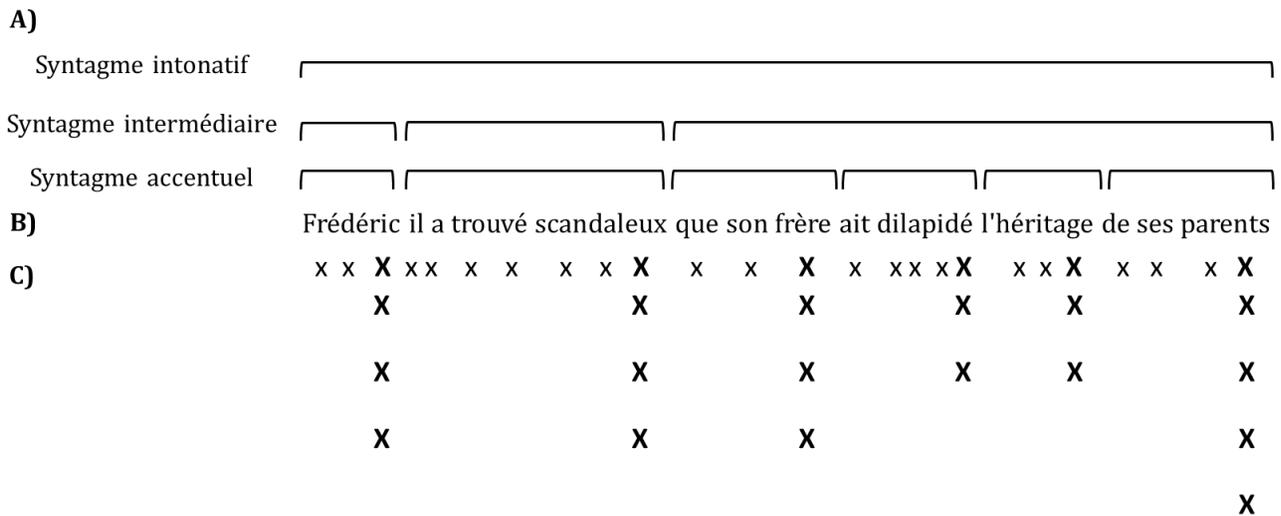


Figure 1.4. Représentation de la structure rythmique ou grille métrique de l'énoncé adaptée de Di Cristo (2003). Regroupement des syllabes en patterns de syllabes accentuées et non accentuées selon la hiérarchie des constituants prosodiques du français et émergence d'une grille métrique. Panel A) Structuration de l'énoncé selon les constituants prosodiques de Jun et Fougeron (2002). Panel B) Enoncé. Panel C) distribution des accents forts et faibles à travers les p-centers et organisation des patterns d'accents selon une hiérarchie de proéminence correspondant aux constituants prosodiques.

Nous aurions donc, en fonction des langues, une représentation perceptive différente de l'organisation temporelle des accents ou grille métrique qui accentuerait la perception

d'une régularité dans la parole. Selon la DAT, ce processus perceptif complexe, similaire à celui réalisé lors de l'écoute de la musique, serait généré par le couplage de plusieurs oscillateurs avec les évènements saillants du stimulus acoustique à différentes échelles temporelles. Lorsque les différents oscillateurs se trouveraient en phase avec le beat, cela augmenterait la perception de l'accent le plus haut dans la hiérarchie autrement appelée le *downbeat* en musique que l'on pourrait assimiler à l'accent de syntagme intonatif dans la hiérarchie prosodique. La perception d'une métrique, en parole comme en musique pourrait être considérée comme une stratégie cognitive supplémentaire pour structurer le signal acoustique et générer des prédictions sur les évènements à venir.

Pour résumer, la perception du rythme serait commune au traitement de la musique et de la parole et elle comporterait au moins deux niveaux : un 1er niveau, le beat en musique et les syllabes (p-centers) dans la parole, et un 2ème niveau la métrique en musique et l'organisation des accents dans les mots et dans les différents constituants dans la parole. L'association de ces deux processus aurait pour effet d'augmenter notre attention sur différents points du signal et possiblement générer des prédictions quant à l'occurrence des évènements de musique et de parole à différents niveaux.

Dans le chapitre qui va suivre, nous allons voir que l'attention et les prédictions temporelles sont des processus cognitifs qui modulent la qualité de traitement de l'input sensoriel en particulier des stimuli auditifs dans des conditions d'écoute difficiles comme la perception dans le bruit ou les conversations à plusieurs.

2. *L'entraînement* rythmique : ou comment utiliser les propriétés de la temporalité de la parole et du fonctionnement neuronal pour améliorer la perception ?

Dans la partie précédente, nous avons vu que la rythmicité des signaux de musique et de parole, i.e. structure rythmique de surface, beat et métrique, permettait au cerveau de "régulariser" l'input auditif afin d'organiser sa perception ce qui aurait pour conséquence d'orienter l'attention sur les évènements pertinents du signal. Nous allons présenter dans cette nouvelle partie différentes études qui montrent que l'orientation de l'attention sur des

événements qui apparaissent de manière régulière dans la musique comme dans la parole, améliore le traitement de l'information auditive. Nous décrivons ensuite comment les différentes théories tentent d'expliquer les effets de la régularité sur le traitement perceptif.

Dans le domaine musical, on sait que la présentation de notes de musique à des intervalles isochrones améliore la capacité à discriminer différentes hauteurs (Jones, Moynihan, Mackenzie, & Puente, 2002). Lorsque des auditeurs doivent juger de la similarité de la hauteur de deux notes disjointes, si ces notes sont espacées par d'autres notes apparaissant à des intervalles réguliers, les auditeurs obtiennent de meilleures performances comparées à une condition dans laquelle les notes seraient espacées par des intervalles irréguliers. On retrouve le même type d'effet de la régularité sur les temps de réaction et l'encodage des sons au niveau cortical et sous-cortical. Tillmann et Lebrun (2006) ont par exemple montré que les auditeurs diminuaient le temps de réaction lors de tâches de décision de similarité de hauteur des notes de musique lorsque les notes à analyser étaient insérées dans des séquences régulières plutôt qu'irrégulières (Tillmann & Lebrun-Guillaud, 2006). Tierney et Kraus (2013) ont également montré grâce à des mesures électrophysiologiques (potentiels évoqués corticaux et du tronc cérébral) qu'un son musical (onset) présenté en même temps qu'un extrait musical est mieux encodé au niveau cortical (i.e. onde P1 de plus grande amplitude) et sous-cortical (onde V de plus grande amplitude) lorsque ce son est présenté sur le beat plutôt que lorsqu'il est présenté en dehors du beat. A travers des tâches de discrimination phonémique ou de temps de réaction, des expériences ont également montré l'effet de la régularité de la parole sur le traitement perceptif (Tierney & Kraus, 2013). L'hypothèse de ces études était, comme en musique, que la régularité d'apparition des événements à différents niveaux de hiérarchie dans le signal permettrait d'orienter, par anticipation, l'attention de l'auditeur sur les points à discriminer et augmenterait les ressources allouées à leur traitement. Quené et Port (2005) ont par exemple montré que lorsqu'on présente auditivement une liste de mots bisyllabiques, les auditeurs détectent plus rapidement un phonème-cible (e.g. le phonème /t/ dans /gato/) lorsque les p-centers des mots sont espacés par des intervalles réguliers comparé à des intervalles irréguliers (Quené & Port, 2005). Pitt & Samuel (1990) ont quant à eux testé les effets de la régularité générée non pas par les intervalles temporels entre les sons de parole mais par la régularité de la

structure métrique des mots. Ils ont montré que lorsque des auditeurs entendent des listes de mots qui ont un pattern métrique régulier (e.g. avec un accent final systématique sur la dernière syllabe de mots bisyllabiques: patterns iambique), ils parviennent à détecter plus rapidement un phonème-cible comparé à une condition dans laquelle, les mots qui s'enchainent, comme dans une phrase, n'ont pas des patterns métriques identiques (Pitt & Samuel, 1990).

Une autre manière d'analyser les effets de la régularité sur la perception de la parole est la présentation, avant l'écoute d'un stimulus de parole, d'un amorçage de type rythmique reproduisant ou pas, la structure métrique de la parole. Cason et Schön (2012) ont par exemple fait entendre à des auditeurs des amorces rythmiques régulières suivies de non-mots bi et trisyllabiques. Ces amorces contenaient deux types de structures métriques marquées par des accents faibles et des accents forts qui correspondaient ou pas à la métrique des non-mots présentés. La tâche des participants était de décider, en pressant sur un boîtier de réponses, s'ils avaient perçu un phonème-cible dans les non-mots entendus. Le phonème à détecter était soit présenté sur la périodicité du beat, générée par l'amorce métrique (i.e. un accent fort apparaissait toutes les 700 ms), soit en dehors du beat et cela, dans les deux conditions d'amorce métrique (identique ou différente). Les résultats montrent que les auditeurs détectent plus rapidement les phonèmes-cible lorsque les phonèmes sont présentés sur le beat plutôt que off-beat et que ce résultat est encore amélioré lorsque la métrique de l'amorce correspond à la métrique du non-mot (effet d'interaction). En outre, des mesures électrophysiologiques (potentiels évoqués corticaux) montrent que les auditeurs ont bien détecté la violation de l'amorce métrique dans la condition où la métrique du mot n'était pas identique à l'amorce (N100) ainsi que la violation de la régularité de présentation du phonème sur le beat (P300) (Cason & Schön, 2012).

Pour résumer, dans la musique comme dans la parole la régularité de présentation des stimuli améliore les performances comportementales et électrophysiologiques dans des tâches perceptives. Lorsque le contexte est régulier, l'attention des sujets est facilement orientée sur les points saillants du signal car leur occurrence peut être anticipée. En outre, la perception de la métrique augmenterait cet effet en "attirant" par anticipation, les ressources attentionnelles sur les syllabes accentuées à différents niveaux.

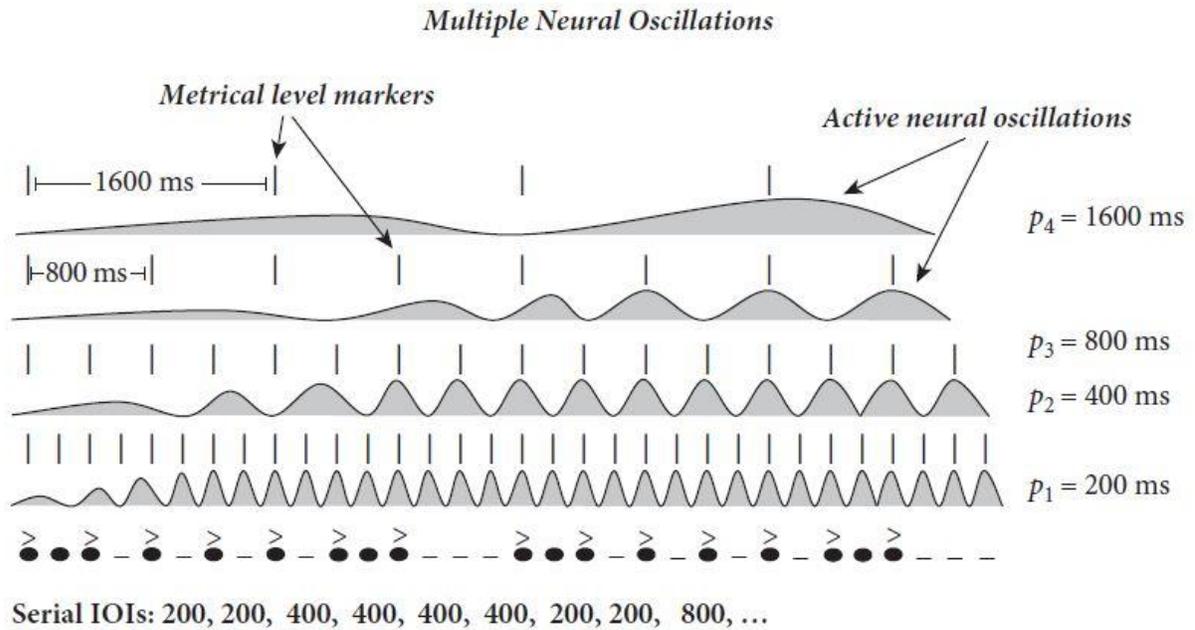
Plusieurs théories ont tenté d'expliquer les processus neurocognitifs qui pourraient être à l'origine de cet effet de facilitation de la régularité sur le traitement perceptif.

Toutes ces théories partent du constat que l'activité du cerveau est de nature rythmique. Lorsque que l'on pratique un enregistrement de l'activité cérébrale (EEG), on peut en effet voir apparaître spontanément et en fonction des opérations cognitives réalisées durant la tâche, des oscillations rythmiques dans différentes bandes de fréquences. On recense essentiellement quatre bandes de fréquences oscillatoires : delta (1-3 Hz), theta (4-7Hz), alpha (8-14Hz), gamma (40-80 Hz). Ces fréquences ont la particularité d'être organisées de manière hiérarchique : les plus lentes "contiennent" les plus rapides. On dit aussi que les fréquences rapides sont "nichées", nested, dans les fréquences plus lentes (figure 1.6). Ces oscillations sont par ailleurs censées refléter les variations cycliques de l'excitabilité d'un ensemble de neurones. Autrement dit, lors du pic d'une oscillation plusieurs neurones se préparent à émettre en même temps, plusieurs potentiels d'action.

Une première théorie propose que selon que les évènements sensoriels arrivent au cortex au moment du pic ou du creux d'une oscillation, leur traitement, du fait du niveau d'excitabilité des neurones, va être respectivement amplifié ou atténué (Schroeder & Lakatos, 2009). Dans cette théorie, les oscillations corticales, par leurs fluctuations régulières, sont considérées comme un moyen pour le cerveau d'échantillonner le flux continu d'informations sensorielles, ou en d'autres termes, un filtre temporel. Plusieurs études dans le domaine visuel, sensorimoteur et auditif, semblent aller dans le sens de cette théorie. Il a par exemple été montré que si la phase des oscillations delta et theta était au niveau d'un pic d'oscillation juste avant l'arrivée d'une syllabe ambiguë, cela permettait aux auditeurs d'améliorer leurs capacités de catégorisation de cette syllabe, autrement dit de mieux la distinguer d'une autre syllabe (ten Oever & Sack, 2015).

Une deuxième théorie, la Dynamic Attending Theory (déjà évoquée dans la 1ère partie), postule que notre système auditif augmenterait le traitement du signal en synchronisant la phase de ses oscillations corticales avec la phase des stimuli rythmiques. La facilitation engendrée par la régularité serait le résultat du couplage automatique de deux oscillateurs, le stimulus rythmique et un oscillateur du cortex auditif dans le cas d'un stimulus

simple comme un métronome ou de plusieurs oscillateurs dans le cas d'un stimulus complexe comme la musique ou la parole (voir figure 1.5). Cette théorie, considère que l'effet facilitateur de la rythmicité dans la musique est un processus de type *bottom-up*, passif, dans lequel les oscillations du système sensoriel seraient entraînées par l'enveloppe temporelle du stimulus acoustique musical. Comme le signal musical, le signal de parole, contient deux types d'informations acoustiques : des informations de nature spectrale et d'autres de nature temporelle. La structure temporelle de la parole peut se décomposer en deux éléments : la structure fine, marquée par des fréquences extrêmement rapides (de 60 Hz à 10 KHz) qui permettent par exemple de différencier deux phonèmes opposés seulement par un trait comme le voisement ou de percevoir les transitions formantiques entre consonnes et voyelles. Le deuxième élément est l'enveloppe temporelle, marquée par des fluctuations lentes de l'amplitude du signal (de 2 à 50 Hz), qui encodent des informations de durée et d'intensité ou encore d'attaque et d'amortissement et qui permettent par exemple de distinguer deux consonnes qui s'opposent par leur mode articulaire (constrictif versus occlusif). L'enveloppe temporelle permet aussi de percevoir le débit syllabique ou encore le rythme de la parole, marqué par la fréquence d'apparition des phonèmes, des syllabes et des accents (Rosen, 1992). Le rythme de la parole, serait ainsi reflété dans les fluctuations lentes de l'amplitude du signal acoustique qui se décomposent elles aussi en différentes bandes fréquentielles contenues les unes dans les autres comme pour les oscillations corticales (voir figure 1.6). On retrouve ainsi le rythme delta (approximativement entre 0.5 et 3 Hz) qui est la fréquence d'apparition des accents des différents constituants prosodiques, le rythme theta (compris entre 4 et 8 Hz) qui est la fréquence d'apparition des syllabes, et le rythme gamma (entre 30 et 80 Hz) qui est la fréquence d'apparition des phonèmes (Greenberg, 1999). Ainsi, l'enveloppe de la parole, par les informations temporelles qu'elle comporte, transmet un nombre d'informations important à différentes échelles temporelles qui vont être utilisées par le cerveau pour segmenter puis assembler les informations acoustiques de manière pertinente. On pourrait ainsi faire l'hypothèse que la régularité d'apparition des stimuli ou les amorces rythmiques délivrées avant le matériel verbal, auraient favorisé la synchronisation des oscillations neuronales du cortex auditif avec l'enveloppe temporelle de la musique et de la parole. Grâce à ce phénomène d'*entraînement*, la perception des informations acoustiques fines aurait pu être améliorée.



A Metric Cluster

Figure 1.5. Représentation du couplage de phase de plusieurs oscillateurs avec les différents niveaux de régularité dans le stimulus acoustique extraite de Jones (2016). Plusieurs oscillateurs avec chacun différentes périodes (P_1 - P_2 - P_3 - P_4), oscillent en même temps à différentes échelles temporelles correspondant à différents niveaux de régularités du stimulus. Lorsque plusieurs oscillateurs entrent en cohérence de phase, ils créent une relation entre les différents niveaux de la hiérarchie métrique : « *metric clusters* ». Une fois les clusters métriques établis, l'attention peut être plus flexible et se focaliser sur différents niveaux dans la hiérarchie et anticiper les événements acoustiques à différents niveaux temporels.

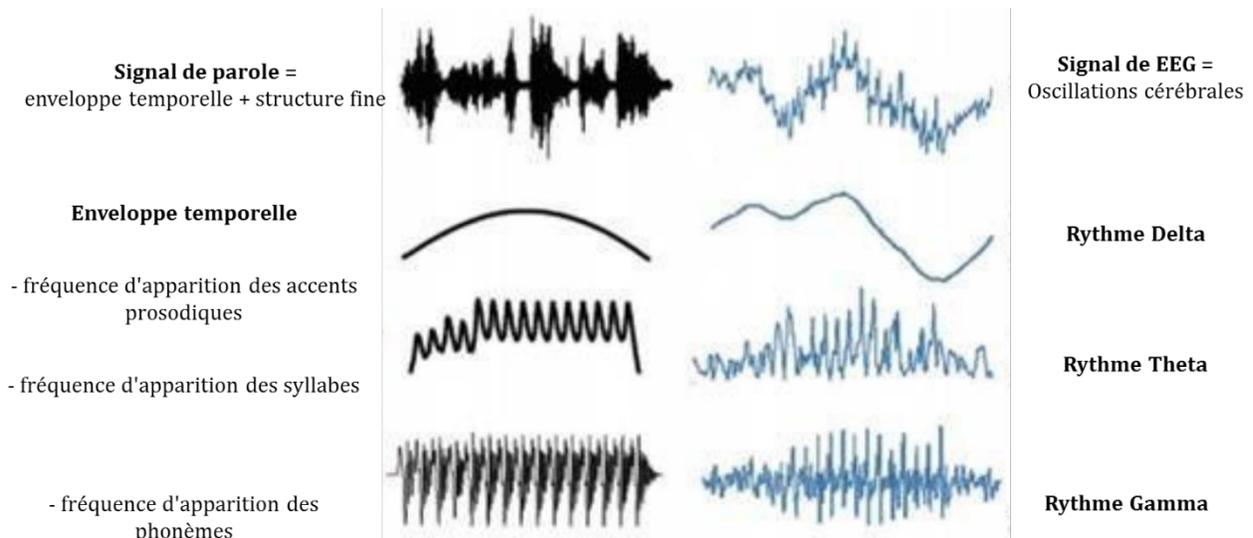


Figure 1.6. Représentation des fréquences dans l'enveloppe temporelle de la parole et dans l'activité cérébrale. A gauche, représentation du signal acoustique et d'une décomposition de l'enveloppe temporelle dans ses différentes bandes de fréquences. A droite, représentation des différentes bandes de fréquences contenues dans le signal EEG et leur caractère "nested" : les fréquences d'oscillation les plus lentes i.e. delta, contiennent les fréquences un peu plus rapides i.e. theta, qui contiennent les fréquences gamma qui sont les plus rapides.

Une troisième théorie la *Predictive Coding theory* propose que l'alignement de la phase (du pic) des oscillations neuronales avec la régularité des stimuli ne soit pas seulement le fait d'un *entrainement* passif de type *bottom-up* mais aussi le résultat d'un mécanisme plus actif de prédictions temporelles. Selon cette théorie, la qualité de l'*entrainement* pourrait être modulée par des processus de type *top-down*. Une manière de mesurer les effets de ce processus *top-down* sur la qualité de la perception est de manipuler les attentes temporelles à plusieurs niveaux en l'absence d'une stricte régularité d'apparition (ou isochronie) des stimuli dans le signal acoustique. Dans une récente étude, Breska & Deouell (2016) ont par exemple manipulé la probabilité d'apparition d'un flash visuel : la cible visuelle pouvait soit apparaître sur le beat (*on-beat*), c'est à dire à un intervalle temporel identique aux intervalles séparant les flashes précédents, soit en dehors du beat (*off-beat*) (Breska & Deouell, 2016).

Cette dernière condition signifie que la cible avait une plus grande probabilité d'apparaître à un intervalle temporel différent de ceux précédemment présentés. Les résultats montrent que lorsque les cibles arrivent à des moments non attendus, les participants mettent plus de temps à répondre comparé à une condition dans laquelle les cibles arrivent de manière aléatoire. Ce qui est intéressant c'est que ce résultat est valable pour la condition *on-beat* comme *off-beat*. En effet, si le résultat obtenu dans la condition *on-beat* peut être imputable à l'*entrainment* du cortex visuel sur la régularité du stimulus, le résultat obtenu dans la condition *off-beat* serait lui, le résultat d'une synchronisation de phase due à la capacité du système à prédire l'occurrence des stimuli selon leurs probabilités d'apparition, et non en fonction de l'isochronie de leur occurrence. Ce type de prédictions, plus globales, faciliteraient également le traitement de l'information sensorielle en orientant les ressources attentionnelles sur des éléments pertinents pour la tâche. De nouveaux paradigmes sont actuellement en construction dans le domaine auditif afin de mesurer les effets de ces prédictions top-down dans le domaine auditif (c.f. Haegens & Zion Golumbic, 2018).

Pour résumer, le fonctionnement oscillatoire de notre cerveau nous permettrait de traiter les stimuli rythmiques avec plus de facilité en définissant des fenêtres attentionnelles de nature temporelles dans le flux de musique et de parole et en synchronisant la phase de ses oscillations sur les événements prédictibles. L'orientation des ressources attentionnelles sur le stimulus acoustique pourrait être le fruit d'un mécanisme d'*entrainment* passif (*bottom-up*) mais qui pourrait être modulé par un autre mécanisme plus stratégique, de prédictions temporelles (*top-down*) à un niveau global. Les oscillations neuronales étant "emboîtées les unes dans les autres", on peut penser que les prédictions plus globales, de type top-down, favoriseraient la synchronisation sur les fréquences lentes de la parole (informations sur la métrique prosodique) qui pourraient "contraindre" les oscillations de fréquences plus rapides à se mettre en phase avec les événements de parole tels que les syllabes et les phonèmes; l'auditeur qui anticiperait la fin d'une unité prosodique pourrait en effet prédire le déroulement temporel des syllabes contenues dans cette unité prosodique. Ces deux processus neurocognitifs, *bottom-up* et *top-down*, de nature temporelle optimiseraient le traitement de la parole aux différentes échelles temporelles et seraient spécifiquement actifs lors de la perception dans le bruit ou de conversations à plusieurs. Ces mécanismes

seraient par ailleurs particulièrement développés lors de la pratique de la musique car cette activité cognitive de nature sensorimotrice, sollicite de manière accrue les compétences de traitement temporel à différents niveaux et nécessite une grande précision temporelle.

Dans la partie suivante, nous allons voir que la pratique de la musique améliore la perception de la parole, probablement grâce à une sollicitation accrue de la communication entre les structures auditives et motrices ce qui permettrait un traitement temporel de la parole plus précis de nature prédictive.

3. La pratique musicale : une activité rythmique qui améliore la perception de la parole

Les études mesurant les réponses électrophysiologiques à des stimuli de parole ont pu montrer que chez les musiciens, les neurones du tronc cérébral (colliculus inférieur), dévolus au traitement précoce de l'information auditive, réagissent plus rapidement et reproduisent de manière plus précise les caractéristiques acoustiques du stimulus (Kraus & Chandrasekaran, 2010; Wong, Skoe, Russo, Dees, & Kraus, 2007). Cela leur permet par exemple de mieux traiter et comprendre la parole dans le bruit (Parbery-Clark, Strait, & Kraus, 2011; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Zendel, Tremblay, Belleville, & Peretz, 2015), de traiter avec plus de précision les différences entre les sons du langage (formants vocaliques pour les voyelles et transitions rapides pour les consonnes) (Parbery-Clark, Tierney, Strait, & Kraus, 2012) ou encore de mieux discriminer les changements de hauteur (Besson, Schön, Moreno, Santos, & Magne, 2007; Schön, Magne, & Besson, 2004) intervenant dans la structure de la prosodie et la valeur émotionnelle contenue dans un message verbal. Il est possible que le meilleur traitement de la parole chez les musiciens provienne de leur capacité à traiter l'organisation temporelle des sons avec plus de précision. Les musiciens discriminent des intervalles temporels plus petits comparé aux non musiciens ou des changements temporels plus subtils à l'intérieur de structures rythmiques (Rammsayer & Altenmüller, 2006). Les musiciens présentent également une réponse électrophysiologique (Négarivité de Discordance : MMN) plus importante dans le cortex auditif gauche que les non musiciens lors de la détection de violations de la métrique en musique (Vuust et al., 2005), ce qui pourrait expliquer qu'ils parviennent à porter une attention plus importante aux

structures métriques dans la musique (Kung, Tzeng, Hung, & Wu, 2011) comme dans la parole. Les musiciens parviennent en effet à mieux détecter une incongruité dans la structure métrique des mots terminant des phrases et leurs réponses électrophysiologiques (P200) à cette incongruité sont plus importantes que celles des non musiciens lorsque la détection de l'incongruité nécessite un traitement actif dans la tâche (i.e. la compréhension des mots) (Marie, Magne, & Besson, 2011). Les enfants musiciens sont également plus performants que les enfants non musiciens à détecter des violations syntaxiques en langage (Jentschke & Koelsch, 2009). Ces résultats semblent signifier que la pratique facilite la perception des structures temporelles contenues dans la musique et dans le langage ce qui influencerait leurs capacités de prédictions dans la parole. Les capacités de prédiction joueraient un rôle important dans l'extraction de régularités statistiques qui sont à la base de la segmentation de la parole. Les musiciens sembleraient lorsqu'il s'agit de segmenter un flux de parole dans une nouvelle langue (Francois & Schön, 2011). Une étude de Vuust et collaborateurs (2008) en magnétoencéphalographie a également montré qu'au plus les musiciens possèdent une expérience rythmique importante (i.e. les musiciens de jazz), au plus leur cortex auditif est sensible à des violations dans les structures rythmiques (MMNm) ; cet effet pourrait être expliqué par leurs capacités à générer des prédictions temporelles de haut niveau reflétées par l'amplitude de la P300m lors des violations métriques (Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2008). Ces prédictions de type *predictive coding*, implémentées dans les cortex frontal et pariétal, viendraient moduler de manière plus importante, comparé à des musiciens moins expérimentés en pratique rythmique ou des non musiciens, la réponse du cortex auditif à bas niveau. Il est intéressant de noter que la pratique du rythme mais aussi son écoute passive, sollicitent le système moteur (i.e. Aire Motrice Supplémentaire : AMS et cortex prémoteur) (J. L. Chen, Penhune, & Zatorre, 2008). Or, l'activité des structures motrices et les structures frontales semblent liées à la prédictibilité des stimuli. Ainsi, Bengtsson et collaborateurs (2009) ont montré dans une tâche qui n'impliquait aucune action motrice, que l'AMS était plus activée lors de l'écoute de structures de tons organisés selon une structure rythmique comparé à des séquences de tons distribués de façon aléatoire (Bengtsson et al., 2009). Les auteurs (Bengtsson et al., 2009) ont également montré que le cortex préfrontal était plus activé lors de l'écoute de structures métriques et non métriques impliquant une complexité rythmique et des capacités de prédiction plus importantes que le

traitement d'une séquence de tons isochrones. Par ailleurs, comme le suggèrent des études montrant l'influence du système moteur dans des tâches de discrimination auditive de hauteur (Morillon, Schroeder, & Wyart, 2014) ou de phonèmes (Cason, Astésano, & Schön, 2015), l'implication du système moteur dans des tâches de perception améliore les performances des auditeurs. Il semblerait donc que la pratique de la musique et en particulier du rythme sollicitant la communication entre les structures motrices et frontales d'une part et les aires auditives d'autres part, soit à l'origine du meilleur traitement des stimuli acoustiques chez les musiciens. L'activation du système moteur lors de l'écoute des sons, augmenterait la synchronisation des oscillations du système auditif sur l'enveloppe des stimuli acoustiques de musique et de parole, ce qui faciliterait l'orientation de l'attention sur le signal acoustique et affinerait leur représentation.

II. L'interaction sociale : un contexte ou le rythme (de parole) est essentiel

Si la rythmicité contenue dans la parole, et de surcroît dans la musique, permet à l'auditeur d'activer des mécanismes neurocognitifs permettant d'améliorer le traitement du stimulus auditif, la situation d'interaction, qui est la situation la plus couramment rencontrée pour le traitement de la parole, va induire des modulations de ces processus (Dimitrios Kourtis, Sebanz, & Knoblich, 2010) et nécessiter des capacités d'anticipation et d'adaptation supplémentaires à celle d'une perception de la parole en situation isolée (Keller, Novembre, & Hove, 2014; Knoblich & Jordan, 2003). C'est pour cette raison qu'il apparaît important de remplacer les mécanismes de perception et de production de la parole dans un contexte interactionnel.

1. Le tour de parole, un contexte interactionnel qui nécessite prédiction temporelle et coordination

Aider un ami à transporter un canapé, jouer au tennis, au basket, danser, ou encore jouer de la musique avec d'autres musiciens requiert une coordination entre les individus pour que l'action puisse aboutir. Les patineurs artistiques doivent par exemple, lors d'une portée, coordonner précisément spatialement et temporellement leurs mouvements afin que

cette figure soit réussie. Cependant, cette coordination inter-individuelle nécessite une constante adaptation aux comportements du partenaire qui peuvent, pour certains s'avérer imprévisibles. Par exemple, lors d'un match de handball, pour qu'une passe soit réussie, le receveur doit anticiper la trajectoire et la vitesse du ballon en fonction des mouvements et déplacements de son co-équipier. Mais il doit aussi tenir compte des différents adversaires qui peuvent faire obstacle au lanceur et provoquer probablement une déviation impromptue de son geste de lancer. Ce phénomène, nommé "coordination interpersonnelle", revêt un caractère dynamique car les individus sont à la fois en constante adaptation mutuelle mais aussi en adaptation avec leur environnement. La coordination qui est présente dans la plupart des interactions sociales, est également présente dans les interactions verbales telles que la conversation qui nécessitent, de la part des interlocuteurs une adaptation ou accommodation temporelle fortement basée sur le rythme. En ce qui nous concerne, nous nous intéressons précisément aux mécanismes neurocognitifs permettant aux interlocuteurs de se coordonner mutuellement pour déterminer avec précision le moment de prise de tour de parole malgré toute la variabilité existante en termes de contenu et de longueur entre deux énoncés provenant de deux interlocuteurs différents.

a. La composante temporelle du tour de parole

Selon l'analyse conversationnelle, les interactions sociales sont régies par un certain nombre de processus et le tour de parole est un des processus majeurs mis en place par les interlocuteurs de manière récurrente dans les différents contextes conversationnels. Un des premiers modèles du tour de parole a été développé par Sacks et al., (1974) dans lequel les auteurs considèrent le tour de parole comme une unité d'interaction sociale construite et réinventée par les interlocuteurs tout au long de la conversation. Deux composantes déterminent un tour de parole : la 1ère composante est définie par les unités verbales c'est-à-dire les mots et phrases utilisés par chaque interlocuteur ; ces unités peuvent être de longueur très différente d'un tour à l'autre et sont appelées *Turn Constructional Unit* (TCU). La seconde composante est définie comme le moment le plus pertinent pour le changement de locuteur ; elle est nommée le *Turn Relevant Place* (TRP) (Sacks, Schegloff, & Jefferson, 1974). Dans ce modèle, plusieurs règles comportementales sont définies et doivent nécessairement être mutuellement respectées et comprises par les interlocuteurs afin que la

prise de tour de parole puisse avoir lieu à un moment jugé opportun par l'un comme l'autre des deux interlocuteurs. Nous ne développerons pas les comportements non verbaux, ni les traitements lexico-syntaxiques qui sont des indices donnés et utilisés par les interlocuteurs pour déterminer la prise ou non du tour de parole. Cependant, l'analyse de l'organisation temporelle des tours de parole dans la conversation révèle que la plupart du temps, les interlocuteurs parviennent à se coordonner précisément afin que la parole de l'un ne chevauche pas celle de l'autre ni qu'il s'écoule un temps de silence trop important entre deux tours. Les auteurs (Sacks et al., 1974) résument ce principe par *minimal gap* et *minimal overlap*.

Selon Auer et collaborateurs (1999), l'organisation temporelle des tours de parole renfermerait un caractère rythmique dont la fonction serait de coordonner l'interaction verbale. Cette rythmicité des échanges permettrait de déterminer si un énoncé s'intègre ou non à un échange conversationnel. Une des stratégies adoptées par les locuteurs (anglophones) pour demeurer coordonnés temporellement lors d'une conversation, serait de modifier certaines caractéristiques de leurs syllabes afin de créer une continuité rythmique entre les tours. En se basant sur la perception des intervalles temporels qui séparent les syllabes accentuées de fin de tour, le locuteur qui se prépare à prendre la parole aura tendance à fixer la durée de la première syllabe accentuée de son tour de parole de manière à créer un intervalle isochrone entre la syllabe de fin tour de son interlocuteur et celle du début de son tour de parole (Auer, Couper-Kuhlen, & Müller, 1999). Une autre stratégie utilisée par les interlocuteurs pour conserver un degré de rythmicité dans l'échange serait la tendance à produire des mots comportant un nombre de syllabes identique d'un tour à l'autre ce qui crée selon les auteurs, une isométrie intra et inter-tour renforçant la perception rythmique de la conversation. Néanmoins, même si les portions de parole contenant une rythmicité forte (mesurées acoustiquement) sont rares, elles sont plus particulièrement présentes vers la fin des tours de parole (Szczeppek Reed, 2010a). Le renforcement local (i.e. aux frontières des tours) des indices rythmiques dans la conversation semble indiquer que la conservation de la rythmicité de la parole joue bien un rôle fonctionnel dans la coordination des tours de parole. Dans une étude Szczeppek Reed (2010b) a analysé au cours d'interviews, la rythmicité des échanges de paroles entre des locuteurs de

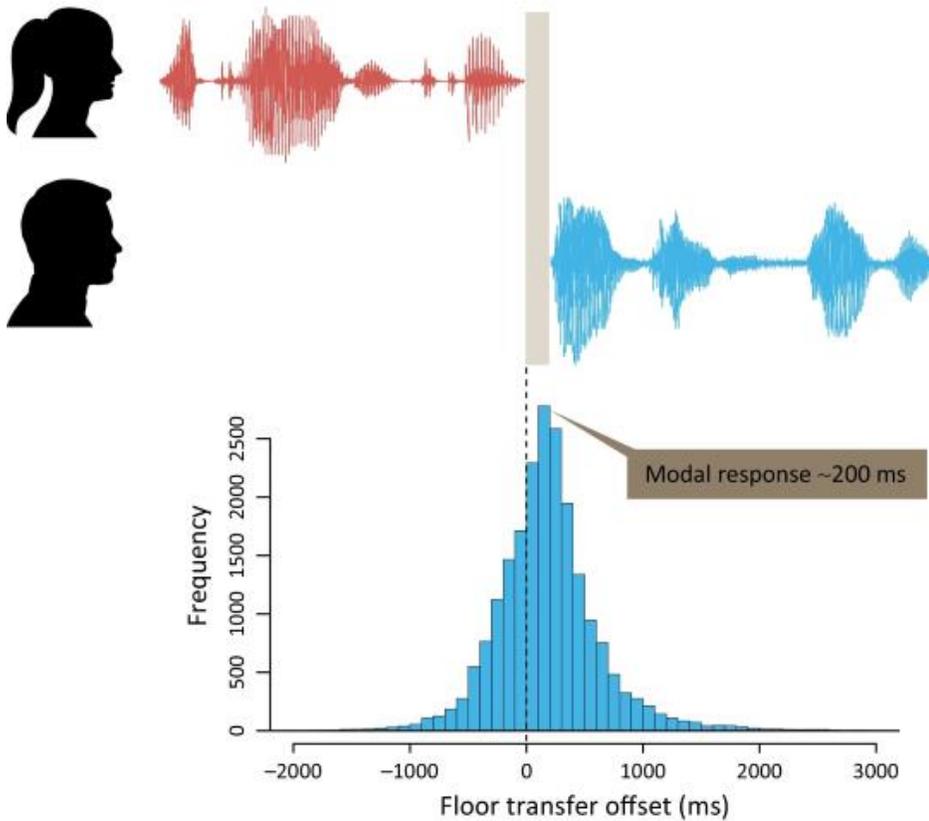
deux communautés anglophones différentes. Ce type de conversation, composé d'échanges de question-réponse, a été spécifiquement choisi pour permettre aux auteurs de contrôler le temps de latence dû à la difficulté d'accès lexical. Les deux communautés participant à ces interviews différaient par la rythmicité de l'anglais parlé. En effet, le British English (BE) et le Singapour English (SE) présentent un rythme de parole différent : le rythme du BE qualifié de *stress-timed*, se reflèterait dans la régularité de la distribution de ses accents et le rythme du SE qualifié de *syllable-timed*, se reflèterait dans la durée isochrone des syllabes (Szczeppek Reed, 2010b). Les résultats montrent que pour ces deux communautés, possédant un rythme de parole différent, la conservation de la rythmicité entre les tours est peu présente ce qui peut altérer la fluidité des échanges et modifier la portée pragmatique de la conversation (Beňuš, Gravano, & Hirschberg, 2011). En effet, si la rythmicité des échanges entre les locuteurs est mise à mal par des temps de latence trop importants ou des chevauchements trop nombreux entre les tours, soit l'interprétation globale des échanges de parole comme activité conversationnelle conjointe est rompue, soit, lorsque cette rupture de la rythmicité apparaît de manière sporadique, elle permet de guider les inférences conversationnelles. Par exemple, un silence d'une durée comprise entre 700 et 800 millisecondes entre deux tours de parole va laisser place à la génération d'une inférence de la part de locuteur. Ce dernier va interpréter le message à venir de son interlocuteur d'une manière différente, (probablement comme une forme de désaccord) de l'interprétation qu'il aurait faite si la réponse avait été articulée selon le délai "conventionnel" compris entre 0 et 200 ms (Kendrick & Torreira, 2015).

b. Les prédictions temporelles dans le tour de parole

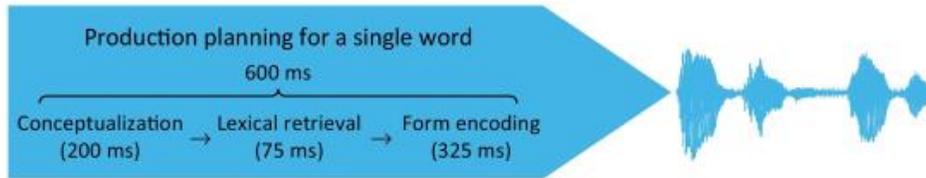
L'analyse temporelle de la prise de tour de parole lors d'un dialogue, montre que le moment auquel les interlocuteurs prennent la parole n'est pas aléatoire. Une étude de Stivers et collaborateurs (2009) a étudié des conversations spontanées dans 10 langues typologiquement éloignées les unes des autres et a pu mettre en évidence une véritable organisation temporelle des tours de parole malgré une variabilité importante dans la structure syntaxique des langues étudiées. Les auteurs ont prélevé les séquences de questions-réponses fermées (oui-non) et mesuré subjectivement, en demandant à un jury d'écoute, si les réponses semblaient retardées ou pas compte-tenu du rythme de la

conversation. Ils ont également mesuré objectivement le temps de réponse en millisecondes par rapport à la fin de la question. Ce temps est positif quand il existe un silence entre la fin de la question et le début de la réponse et négatif quand il y a un chevauchement de parole. Lors de l'analyse de cet intervalle temporel dans les 10 langues différentes, on peut voir apparaître un pic unimodal s'étendant en moyenne de 0 à 200 ms (Stivers et al., 2009)(voir aussi l'avis nuancé sur la précision temporelle du tour de parole de Heldner & Edlund, 2010). Autrement dit, les locuteurs appartenant à différentes cultures et parlant différentes langues appréhendent une unité temporelle commune lors de leurs échanges verbaux ; les auteurs parlent d'une organisation universelle du tour de parole. La capacité des locuteurs à respecter ce temps "commun" entre les tours nécessite des capacités de prédiction temporelles. En effet, comme l'explique Levinson et Torreira (2015), la production d'un mot prend en moyenne 600 ms. Or, si les interlocuteurs parviennent à produire leur parole dans les 200 ms qui suivent la fin du tour de parole précédent, c'est vraisemblablement qu'ils ont anticipé la fin de la parole du locuteur. La distribution du temps écoulé entre la fin du tour de parole du locuteur précédent et la prise de parole du locuteur suivant est nommée le *Floor transfer offset* (Levinson & Torreira, 2015) (voir figure 2.1). Lorsqu'on demande à des adultes de déterminer le moment le plus probable de la fin des plusieurs tours de parole issus de conversations spontanées, ils estiment ce moment (en pressant un boîtier de réponse) en moyenne 200 ms avant la fin des tours. La manipulation des indices acoustiques a également permis aux auteurs de conclure que cette anticipation était basée sur des indices lexicosyntaxiques (De Ruiter, Mitterer, & Enfield, 2006; Magyari & de Ruiter, 2012; voir Riest, Jorschick, & de Ruiter, 2015 pour une anticipation à partir des indices sémantiques).

(A) Responses in conversation are fast



(B) Latencies in production are threefold or more longer than the modal gap



(C) Production of response must therefore overlap with comprehension of the incoming turn

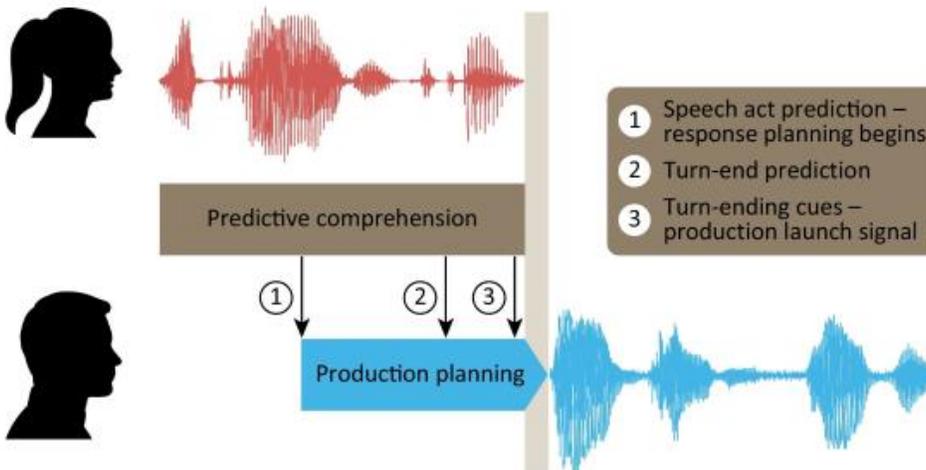


Figure 2.1. Représentation du déroulement temporel des tours de parole extraite de Levinson (2016). A) La parole du locuteur A (en rouge) se termine au point 0 du *Floor Transfert Offset* marqué par une ligne en pointillée. La durée des silences laissée par les interlocuteurs entre les tours est ainsi calculée depuis ce point 0 jusqu'au début de la parole du locuteur B (en bleu). La médiane des silences inter-tours se situe autour de 200 ms. B) Lorsqu'un locuteur produit un mot, les différentes étapes qui vont de la conceptualisation au début de la réalisation motrice prennent 600 ms. C) Pendant que locuteur A produit son message, le locuteur B génère constamment des prédictions en termes de contenu et de déroulement temporel du message. Ainsi en 1 le locuteur B prédit le type d'acte de parole (e.g. affirmation, question) que va produire le locuteur A en fonction du contenu précédent. En 2 le locuteur prédit le moment de la fin du tour de parole et en 3, en se basant sur les indices acoustiques de la fin du tour de parole, il détermine le moment où il va pouvoir commencer à parler. Pendant ces trois différentes étapes 1-2-3, le locuteur B peut ainsi planifier sa réponse et la produire entre 0 et 200 ms en moyenne malgré le temps de planification-programmation que nécessite la production d'un mot (Levinson, 2016).

On peut se demander par quels mécanismes neurophysiologiques et processus cognitifs les locuteurs parviennent à anticiper la fin des tours de parole et à s'adapter aux variations temporelles présentes lors d'une conversation. Selon Keller, Novembre et Hove (2014), il est possible d'appréhender les compétences de coordination dans l'interaction verbale à la lumière des compétences requises lors de n'importe quelle action conjointe de nature rythmique car elles requièrent toutes un degré de précision de l'ordre de dizaines de milliseconde et des facultés d'accommodation temporelles. Ce type d'action conjointe, telle que la conversation ou la musique d'ensemble, engagerait des facteurs de réussite communs. Plusieurs facteurs entreraient en jeu lors de la réalisation d'une action conjointe tels que des capacités cognitives et sensori-motrices comme l'anticipation, l'adaptation et l'attention. Ces capacités seraient cependant affectées par d'autres facteurs tels que les connaissances communes des individus sur les règles du système, le degré de familiarité qu'ils entretiennent (e.g. le style de musique ou le degré de formalité de la conversation), le but de l'action et la stratégie utilisée pour réaliser l'action. Des facteurs sociaux tels que la personnalité affecteraient également la réussite de l'interaction (P. E. Keller et al., 2014) (voir figure 2.2).

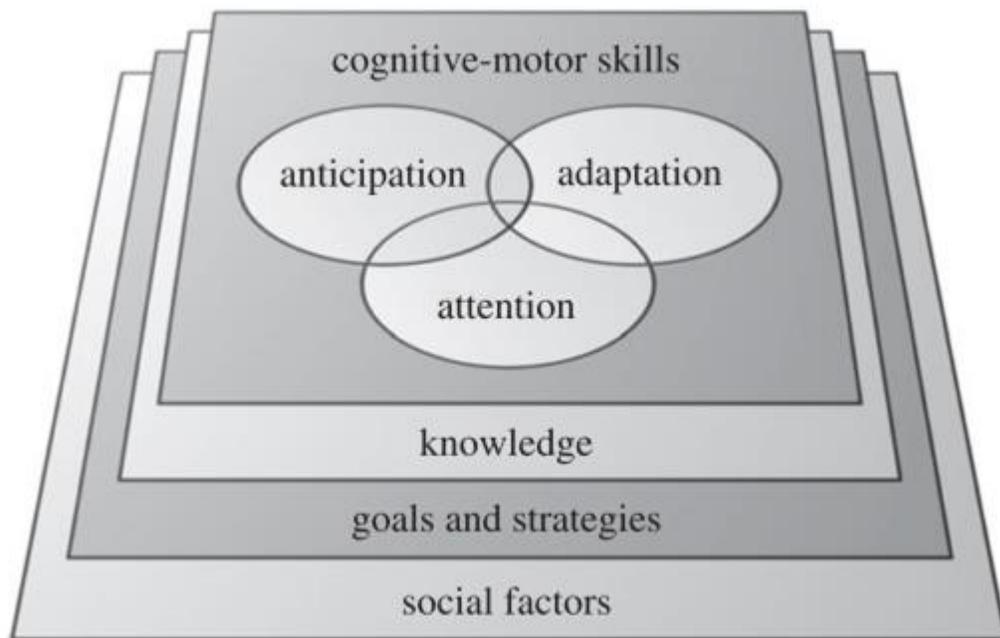


Figure 2.2. Représentation des facteurs qui affectent la coordination interpersonnelle pendant une action conjointe de type rythmique extraite de Keller et collaborateurs (2014).

2. Les mécanismes d'anticipation et d'adaptation dans l'interaction sociale

a. La convergence

L'interaction sociale engendre naturellement chez les individus de véritables comportements d'imitation réciproque. C'est ce que Chartrand & Bargh (1999) nomment l'effet caméléon. Cet effet est observable au niveau des postures, des mimiques faciales ou encore de certaines caractéristiques linguistiques. Chartrand & Bargh (1999) montrent par exemple que des inconnus qui se rencontrent dans le but de réaliser un travail commun vont automatiquement battre du pied ou se toucher le visage si leur partenaire de travail présente un de ces deux comportements. Ce type d'imitation est nommé behavioral matching (Chartrand & Bargh, 1999).

Les interlocuteurs en situation d'interaction verbale modifient également leur manière de parler suite à l'influence exercée par leur partenaire conversationnel. Les interlocuteurs ont spontanément tendance à imiter réciproquement certaines

caractéristiques de leur parole et selon la Théorie de l'Accommodation et de la Communication (Giles, Coupland, & Coupland, 1991) ces comportements imitatifs auraient pour but d'obtenir un maximum d'affiliation, de solidarité entre les locuteurs ainsi qu'une communication optimale i.e. une communication où les interlocuteurs parviennent à établir une inter-compréhension la plus fine possible (Fusaroli & Tylén, 2012).

Les interlocuteurs vont par exemple utiliser les mêmes structures linguistiques (Branigan, Pickering, & Cleland, 2000) et ce, même lorsqu'ils ne possèdent la même langue (Hartsuiker, Pickering, & Veltkamp, 2004), ils vont imiter des paramètres acoustico-phonétiques tels que l'intensité vocale (Natale, 1975), l'accent (Bourhis & Giles, 1977), une combinaison de caractères acoustiques (durée, fréquence fondamentale, les deux 1er formants, Pardo, 2006) mais également des paramètres avec une dimension temporelle tels que le débit de parole (Manson, Bryant, Gervais, & Kline, 2013; Street, 1984).

Ce type d'imitation, nommée convergence, est corrélée, comme le postulait la Théorie de l'Accommodation et de la Communication, au degré d'affiliation et de compréhension entre les interlocuteurs (Branigan et al., 2000; Garrod & Pickering, 2009; Street, 1984). Ainsi, en imitant réciproquement leurs caractéristiques linguistiques, donc en s'alignant sur des caractéristiques de bas niveau, les locuteurs "se rapprocheraient conceptuellement" (voir la notion de *situational models* de Zwaan & Radvansky, 1998), devenant ainsi plus prédictibles et diminueraient l'effort cognitif engagé dans la compréhension de la conversation (Garrod & Pickering, 2004; Pickering, 2006; Pickering & Garrod, 2007).

Levitan et collaborateurs (2015) ont également montré qu'au plus les interlocuteurs convergent sur des paramètres acoustiques à la fin du tour de parole, au plus ils parviennent à conserver des latences similaires entre les tours. Le phénomène de convergence sur la parole pourrait ainsi également servir aux locuteurs à se rapprocher "rythmiquement" dans la conversation afin de mieux se coordonner (Levitan, Beňuš, Gravano, & Hirschberg, 2015).

b. Le phénomène d'*entrainement* sur la parole

L'interaction semble être un contexte propice à la synchronisation temporelle entre les individus. Plusieurs études ont en effet montré que lorsque des individus se rencontrent, ils peuvent adopter naturellement des comportements rythmiques similaires et que ces

comportements sont synchrones, c'est à dire que les deux individus effectuent des mouvements réguliers en même temps comme lorsque deux personnes marchent côte à côte et qu'elles synchronisent leurs enjambées. Une expérience de Richardson et collaborateurs (2007) a par exemple montré que des individus assis côte à côte sur un rocking-chair vont spontanément synchroniser leurs balancements tout comme des individus qui se voient mutuellement en train de faire osciller un pendule, vont naturellement les faire bouger de manière synchrone en phase (0°) ou en antiphase (180°) à la manière de deux oscillateurs qui se couplent. Ce phénomène d'imitation temporelle, commun aux deux études, s'établit alors qu'aucune instruction de coordination n'a été préalablement donnée aux participants (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). Cette forme d'imitation est appelée *interactional synchrony*. En se synchronisant mutuellement, les individus réduiraient leur variabilité intra-individuelle et deviendraient plus prédictibles l'un pour l'autre ; ce phénomène pourrait induire une augmentation de la précision des prédictions temporelles et une facilitation de la coordination inter-individuelle (Vesper, Van Der Wel, Knoblich, & Sebanz, 2011).

L'interactional synchrony, autrement dit la synchronisation spontanée en phase et en anti-phase entre des individus peut se retrouver également lors d'une interaction verbale. Himberg et collaborateurs (2015) ont montré par exemple montré que lorsque deux individus construisent une histoire en produisant des mots chacun leur tour à la manière d'une tâche de tapping en alternance, ils adaptent spontanément leur parole de manière à préserver des intervalles temporels de même durée que ceux réalisés par leur interlocuteur. En effet, même si la longueur des mots est très variable d'un tour à l'autre, les individus s'appuient sur le rythme de parole pour placer le début de chaque mot en fonction de l'intervalle temporel précédent. Cette synchronisation de la parole en anti-phase engendre la perception d'une rythmicité dans l'interaction verbale qui persiste même dans la situation où les interlocuteurs ne peuvent pas se voir (Himberg, Hirvenkari, Mandel, & Hari, 2015). Selon Wilson et Wilson (2005), ce serait la synchronisation mutuelle des oscillateurs internes des interlocuteurs sur le débit syllabique, autrement dit le couplage de deux oscillateurs, qui leur permettrait de prédire et préparer temporellement leur prise de tour de parole afin de rester synchronisés (Wilson & Wilson, 2005).

Cependant, comme le montrent d'autres études, il n'est pas nécessaire que la parole présente une structure périodique parfaite pour que des interlocuteurs parviennent à se synchroniser. Lire un texte ensemble est par exemple une activité très facile pour des interlocuteurs inconnus et inexpérimentés. Il semblerait en effet que même dans des conditions où le débit de parole n'est pas régulier, les individus parviennent à développer à partir du signal, des prédictions temporelles leur permettant de lire les mots en même temps à 40 ms près (Cummins, 2003). Ce résultat semble signifier que la parole, même peu régulière comme lors de la conversation spontanée, peut être un vecteur de synchronisation entre les individus et cette synchronisation inter-individuelle, bénéfique à l'interaction (Valdesolo, Ouyang, & Desteno, 2010), semble solliciter, comme la convergence, la voie sensori-motrice.

c. La voie sensori-motrice et les émulateurs internes

La synchronisation des mouvements des partenaires lors d'une interaction, comme le phénomène de convergence, semblent ressortir d'une synchronisation de type sensori-motrice : le système sensoriel perçoit un stimulus auditif (et/ou visuel) plus ou moins périodique, sur lequel il s'appuie pour envoyer un message au cortex moteur afin d'effectuer une action en fonction des caractéristiques de ce stimulus. Mais le système sensori-moteur fonctionne en boucle, c'est à dire avec des interactions bidirectionnelles entre cortex auditif et cortex moteur. Ainsi, si le système sensoriel interagit avec le système moteur afin qu'il se réajuste suite aux conséquences sensorielles des précédents mouvements, le cortex moteur renvoie également au système sensoriel des informations sur les possibles conséquences sensorielles des mouvements programmés (anticipation). Le système sensoriel peut alors renvoyer des informations au système moteur avant que le mouvement ne soit effectué pour qu'il atteigne la plus grande précision possible. Les conséquences réelles des mouvements, récupérées par le système sensoriel permettront, à force de répétitions, de réduire l'erreur de prédiction. Autrement dit, le système sensori-moteur utilise des mécanismes d'apprentissage par l'erreur qui permettent de réduire facilement la distance entre les mouvements effectués et le but à atteindre (voir figure 2.3 et Pickering & Clark (2014) pour la distinction entre *Auxillary Forward Models* et *Integral Forward Models*).

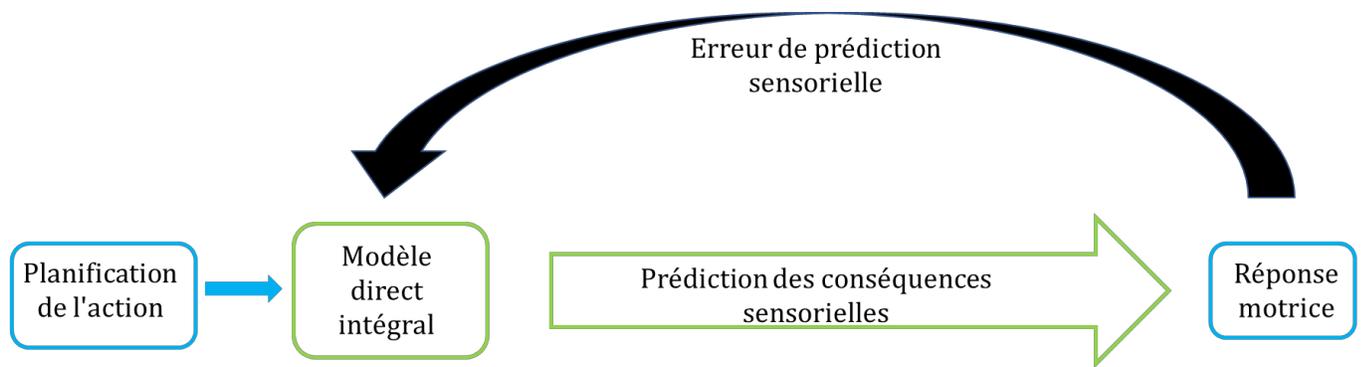


Figure 2.3. Modèle de prédiction de la perception et du réajustement de la réponse motrice dans l'interaction, adapté de Pickering et Clark (2014).

Selon le modèle direct intégral (ou *Integral Forward Model*) de Pickering & Clark (2014), nous utiliserions le même module de prédiction - "modèle direct intégral" - pour anticiper les conséquences sensorielles de nos propres actions et celles de notre interlocuteur mais aussi pour ajuster nos programmes moteurs en fonction des comportements de l'autre. Ainsi, lors de la planification de la parole dans le tour de parole, nous anticiperions les conséquences temporelles de la parole de notre interlocuteur en utilisant les capacités de simulation de notre propre système moteur et nous pourrions ainsi planifier notre parole suffisamment à l'avance afin que le tour commence au moment optimal. Notre modèle serait systématiquement mis à jour et affiné en fonction de la différence entre les conséquences sensorielles prédites par le "modèle direct intégral" et les conséquences sensorielles produites réellement.

Ce sont ces mêmes interactions entre les systèmes perceptifs et le système moteur qui sont à l'œuvre lors de la perception de la parole.

Depuis la découverte du système des neurones miroir (aire F5 du cortex prémoteur ventral) (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), plusieurs études ont montré que lorsque nous percevons de la parole, les régions motrices dévolues à la production de la parole telles que gyrus frontal inférieur gauche, le cortex prémoteur ventral et le cortex moteur primaire s'activent. En outre, on retrouve plus spécifiquement l'activation de régions somato-sensorielles liées aux mouvements de la bouche. Par exemple la partie du cortex moteur dévolue à l'activation des muscles de la langue s'active chez l'auditeur lorsque ce

dernier entend des mots impliquant la mobilisation de la langue d'un locuteur (Fadiga, Craighero, Buccino, & Rizzolatti, 2002). De même, stimuler à l'aide de la TMS les régions somato-sensorielles spécifiques à la réalisation articulaire des certains sons de parole avant la perception de mots contenant ces sons, facilite le traitement en termes de temps de réaction et de pourcentage d'erreurs (D'Ausilio, 2009).

Un modèle de perception de la parole développé par Hickock et Poeppel (2007) met en évidence deux voies de traitement à savoir : une voie dorsale, avec des connexions bidirectionnelles qui relient les régions auditives temporales supérieures aux régions frontales en passant par la jonction temporo-pariétale ; cette voie aurait pour rôle de mettre en correspondance les sons et leurs représentations motrices. Et une voie ventrale qui assurerait la compréhension du message oral (représentations conceptuelle) via des liaisons entre les régions temporales supérieures et inférieures (Hickok & Poeppel, 2007) (voir figure 2.4).

Ainsi lorsque nous percevons de la parole, nous ne percevons pas seulement des éléments acoustico-phonétiques mais, grâce à l'activation des régions motrices, nous intégrons aussi les unités articulatoires correspondantes. En d'autres termes, comme le montre la Perception for Action Control Theory de Schwartz et collaborateurs (2008), nos représentations des unités de parole sont des représentations sensori-motrices (Schwartz, Basirat, Ménard, & Sato, 2012).

Ce modèle rejoint la théorie de la simulation (Wilson & Knoblich, 2005) ou de la résonance motrice qui postule que l'activation de notre système moteur, lors de l'observation de notre partenaire, nous permettrait lors d'une action conjointe, de générer des prédictions sur ses actions à venir et influencerait notre perception. Le système moteur semble en effet très impliqué dans nos interactions sociales, certains auteurs parlent même de "cognition motrice" (Jackson & Decety, 2004). Ce système comporterait un codage commun de l'information avec les systèmes sensoriels et serait le générateur de représentations internes nommées *émulateurs*. Ces représentations motrices de haut niveau étant partagées entre les individus, elles leur permettraient d'anticiper mutuellement leurs actions ainsi que leurs conséquences sensorielles. Cette théorie a été reprise pour la perception de parole (Gambi &

Pickering, 2013) et expliquerait pourquoi l'imitation, qu'elle soit externe comme dans le cas de la convergence ou interne dans le cas d'une imitation motrice endogène des gestes du locuteur de la part l'auditeur, permettrait aux interlocuteurs de mieux se comprendre (Adank, Hagoort, & Bekkering, 2010; Adank, Rueschemeyer, & Bekkering, 2013) et de se coordonner avec précision lorsqu'ils entrent en interaction (Galantucci & Sebanz, 2009).

Scott, Mcgettigan et Eisner (2009) font en effet l'hypothèse d'une activation des cortex moteurs des auditeurs lorsqu'ils sont en train d'écouter un locuteur dans le cadre d'une situation d'interaction verbale. Ils postulent que l'activation de la voie dorsale (reliant les aires sensorielles auditives et les aires motrices en passant par la jonction pariétale) permet, lors du traitement de la parole, une activité motrice finement coordonnée qui assure un déroulement conversationnel fluide. Si la voie ventrale a pour rôle de décoder le message du locuteur d'un point de vue conceptuel, la voie dorsale serait dévolue dans les interactions verbales à contrôler les propriétés temporelles des échanges. Dans la voie dorsale, le rôle du système moteur - système impliqué dans la succession temporelle des actions - serait de suivre, i.e. d'imiter de manière endogène - le rythme et le débit de parole du locuteur et de permettre à l'auditeur d'anticiper la fin des tours de parole (Scott, Mcgettigan, & Eisner, 2009)(voir figure 2.4). Plus précisément, Hadley et collaborateurs (2015) ont montré que dans une tâche lors de laquelle deux pianistes doivent jouer les parties distinctes d'un morceau en alternance (un joue la main la gauche puis l'autre joue la main droite), la perturbation par TMS du cortex prémoteur dorsal et de l'aire motrice supplémentaire au moment de la transition entre les deux pianistes perturbait la fluidité temporelle dans la prise de tour des pianistes. Afin de mesurer l'importance de la simulation des gestes et de la conséquence de cette simulation endogène sur précision de la coordination temporelle dans la tâche, les morceaux ont été préalablement appris soit mains ensemble, autrement dit, pour certains morceaux, les pianistes connaissaient leur propre partie mais aussi celle de leur partenaire, soit les pianistes ont appris seulement la main les concernant. Les résultats montrent que plus la simulation motrice des gestes du partenaire est importante (mesurée pour les morceaux appris mains ensemble), meilleure est la précision de la coordination temporelle lors de la tâche (Hadley, Novembre, Keller, & Pickering, 2015). Novembre et collaborateurs (2014) ont également montré l'importance de la simulation motrice dans

l'adaptation temporelle (coordination). Dans cette étude, des pianistes devaient jouer seulement la main droite de morceaux de musique dont la main gauche était enregistrée ; pour certains des morceaux, les musiciens se sont entraînés et les connaissaient par cœur mais pas pour d'autres. Les pianistes devaient ensuite s'adapter aux changements de tempo imposés par la main gauche enregistrée mais lorsque leur cortex moteur primaire a été perturbé (par d'un système de stimulation magnétique transcrânienne) au moment du changement de tempo, les pianistes ne parvenaient plus à s'adapter aux changements avec autant de précision pour les morceaux qu'ils connaissaient par cœur donc pour lesquels ils étaient capables de générer des simulations motrices. Par ailleurs, de manière intéressante, cette même étude a montré qu'au plus une personne est empathique, au plus l'effet de la perturbation de son système moteur lors de la tâche d'adaptation temporelle est important ce qui suggère que le niveau d'empathie pourrait influencer degré de résonance motrice d'un individu lorsqu'il se trouve en situation d'interaction et pourrait avoir un effet sur ses capacités de coordination (Novembre, Ticini, Schütz-Bosbach, & Keller, 2014). La résonance motrice, impliquant l'Aire Motrice Supplémentaire, le cortex prémoteur dorsal et le cortex moteur primaire semble ainsi indispensable à la génération de prédictions temporelles dans les interactions de type musical mais aussi dans d'autres interactions (Dimitrios Kourtis et al., 2010) comme le tour de parole .

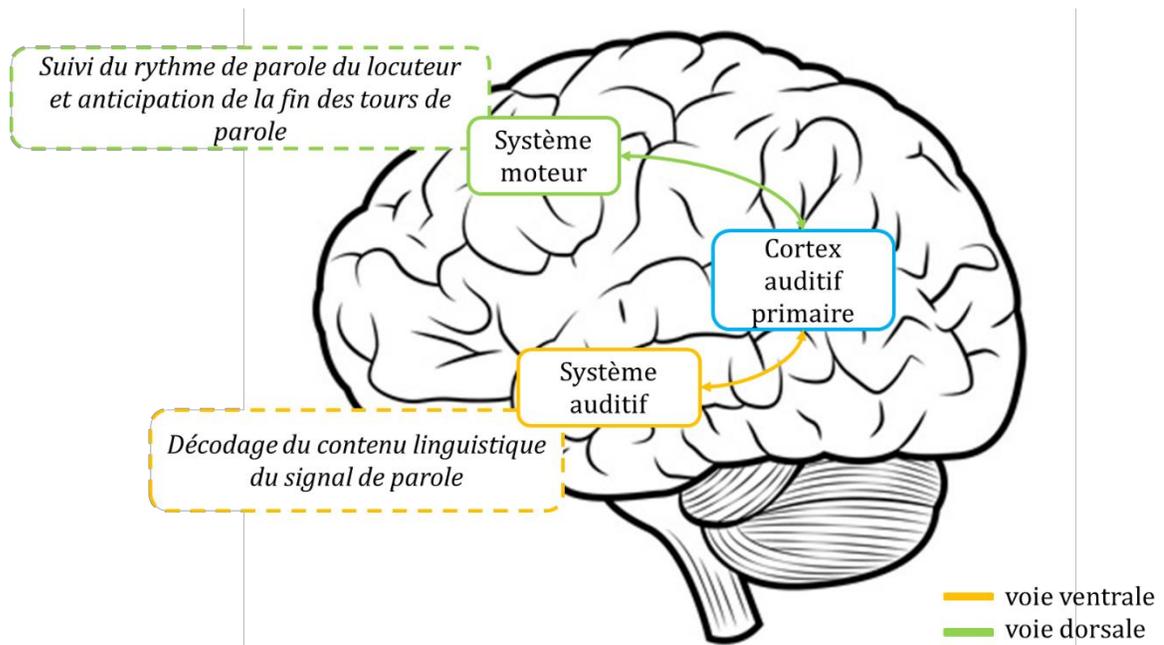


Figure 2.4. Représentation du rôle de la voie ventrale et de la voie dorsale dans le traitement de la parole, adaptée de Scott et collaborateurs (2009).

Les résultats de l'étude de Foti et collaborateurs (2016), convergent en effet vers le modèle de Scott et collaborateurs (2009) postulant l'importance de l'activation endogène du système moteur dans la génération des prédictions temporelles nécessaires au tour de parole. En réalisant un enregistrement EEG simultanément à l'écoute de conversations téléphoniques contenant des paires de questions - réponses car pour ce type de paires, le moment à partir duquel la réponse sera produite est hautement prédictible, ces auteurs ont réussi à montrer que l'anticipation temporelle de la prise des tours de parole avec une durée du silence entre les tours comprise entre 200 et 700 ms (fourchette temporelle qui prédit un accord entre les interlocuteurs et donc une réponse positive immédiate) était reflétée par une réponse évoquée (stimulus-preceding negativity : SPN) générée par des régions postérieures (gyrus temporal postérieur - gyrus supramarginal - cortex prémoteur et le médial cortex préfrontal) correspondant à la voie dorsale alors que le traitement de la réponse affirmative à une question délivrée après une durée de silence supérieure à 700 ms donc engendrant une négation suite à un désaccord entre les interlocuteurs, était reflétée par

le complexe P2/N2/P3 générée par des activations allant du gyrus temporal supérieur vers le médial cortex préfrontal mais n'impliquant pas le cortex moteur. Si l'apparition de ce complexe dans la condition de la réponse non prévisible signe une allocation spécifique de l'attention sur les réponses inattendues, l'élicitation de la SPN, dans la condition attendue suggère l'implication de prédictions temporelles guidées par le système moteur lors de la perception d'une conversation (Foti & Roberts, 2016).

Ainsi, l'imitation et la synchronisation comportementale et verbale avec un partenaire sont des stratégies que développent naturellement les individus lors d'une interaction probablement parce - qu'elles impliquent une activation du système sensori-moteur et en particulier le développement de modèles internes générateurs de prédictions temporelles. La participation importante du cortex moteur dans ces prédictions est une des sources des capacités d'ajustement lors de changement temporels lors d'une interaction.

Un autre processus, complémentaire au phénomène d'entraînement (*bottom-up* et *top-down*) sur un stimulus auditif et du couplage sensori-moteur, pourrait être à l'origine des prédictions temporelles développées lors de l'interaction. Le couplage de phase de l'activité neuronale de deux personnes en train d'interagir serait en effet un processus supplémentaire qui pourrait expliquer les capacités de synchronie interactionnelle développées au cours d'une interaction (voir figure 2.5).

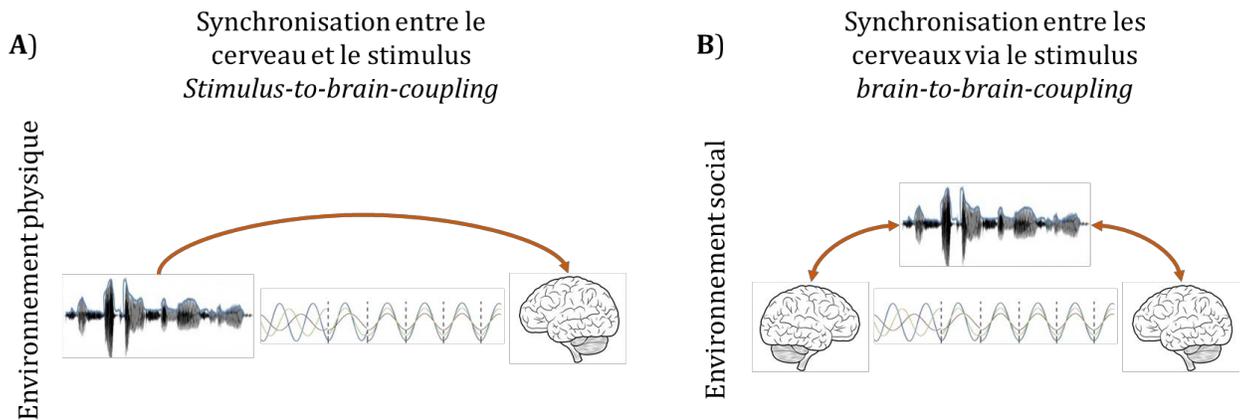


Figure 2.5. Représentation de la synchronisation *stimulus-to-brain* et *brain-to-brain coupling* adaptée de Hasson et collaborateurs (2012). A) Phénomène d’entraînement de l’activité oscillatoire sur un stimulus sensoriel extérieur (*bottom-up*) et de la modulation de l’entraînement par des processus attentionnels et de prédiction (*top-down*). B) Phénomène d’entraînement entre les activités oscillatoires de deux individus différents qui passe par le phénomène vu dans A) c’est à dire par la transmission entre les individus, d’un stimulus qui va leur permettre de synchroniser leurs activités cérébrales.

d. Synchronisation brain to brain

L’adaptation récente des protocoles expérimentaux utilisant l’imagerie, aux situations d’action conjointe, a permis de mieux comprendre le développement des prédictions temporelles durant les interactions en mesurant simultanément l’activité des cerveaux de deux individus.

Cette technique, appelée hyperscanning, a pu mettre en évidence que la synchronisation comportementale serait en quelque sorte “mimée” au niveau cérébral : lorsque deux individus parviennent à se synchroniser pendant une tâche motrice ou verbale, leurs cerveaux présentent également une activité synchronisée sur différentes bandes de fréquences.

Dumas, Nadel, Soussignan, Martinerie et Garnero (2010), ont réalisé des analyses EEG de cerveaux de couples de participants en train d'imiter réciproquement et librement les gestes de leur main. Les analyses révèlent une activation synchrone des régions centro-pariétales droites des deux individus dans les bandes de fréquences alpha-mu. L'activation synchrone de ces régions, particulièrement impliquées dans la perception temporelle, est retrouvée lorsque les participants commençaient et finissaient spontanément leurs mouvements de mains au même moment, autrement dit lorsqu'ils étaient dans des phases de synchronie interactionnelle (Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010).

Une autre étude (Kawasaki, Yamada, Ushiku, Miyauchi, & Yamaguchi, 2013) a utilisé un paradigme expérimental semblable à celui utilisé chez Himberg et collaborateurs (2015) en demandant à des sujets de prononcer les lettres de l'alphabet chacun leur tour tout en enregistrant leur activité cérébrale. Entre chaque série de lettres prononcées en alternance avec un partenaire humain, donc avec une parole plus ou moins irrégulière, chaque individu a effectué le même type de tâche avec une machine programmée pour délivrer chaque lettre à des intervalles temporels réguliers. Après analyse de la durée des productions de chaque participant et des intervalles qui séparent leurs productions (i.e. le rythme de parole et d'alternance) les résultats montrent que les rythmes de parole des participants sont mieux synchronisés dans la situation où les individus alternent avec un humain plutôt qu'avec une machine et que cette synchronisation entre les individus est plus importante après que les individus ont été stimulés par une rythmicité commune délivrée par la machine. Par ailleurs, les analyses EEG montrent une activité oscillatoire synchrone de leurs régions pariétales et temporales dans les bandes de fréquences thêta et alpha. En outre, l'augmentation de la synchronisation sur la parole a renforcé le degré de synchronisation de l'activité neuronale entre les individus (Kawasaki et al., 2013). En situation de conversation spontanée, on peut penser que le cortex auditif de l'auditeur serait entraîné sur les basses fréquences véhiculées dans l'enveloppe temporelle de la parole du locuteur ce qui lui permettrait d'anticiper le moment de la fin du tour de parole de ce dernier (Wilson & Wilson, 2005) et que l'activité synchrone dans des aires impliquées dans la perception temporelle et sur des bandes de fréquences beta dans les aires motrices (Novembre et al., 2017) permettrait aux

interlocuteurs, par un effet prédictif *top-down*, de mieux se coordonner (Garrod & Pickering, 2015).

Ces résultats montrant que la synchronisation de l'activité neuronale entre les individus est liée à la synchronisation sur les rythmes de parole mais uniquement dans une situation d'interaction réelle, c'est à dire où les deux individus sont des humains (i.e. variables et adaptatifs) qui réussissent une tâche interactive (i.e. qui atteignent un but commun ou se comprennent), nous amènent à penser que la synchronisation neuronale interindividuelle joue un rôle prépondérant dans la coordination.

Par ailleurs, il semblerait que l'activité cérébrale d'un auditeur qui anticipe la parole d'un locuteur soit liée à une bonne compréhension du message sonore émis par le locuteur. Une étude (Stephens, Silbert, & Hasson, 2010) en Imagerie par Résonance Magnétique Fonctionnelle a été menée sur un individu en train de raconter une histoire à un autre individu. L'analyse de l'activité spatio-temporelle corticale des deux individus montre des patterns d'activité communs entre les deux individus (narrateur et auditeur) dans les aires auditives primaires lors du récit de l'histoire (i.e. lors de la vocalisation) et une activité anticipée (i.e. qui commence avant la perception de la parole) des aires préfrontales et du striatum chez l'auditeur. Cette synchronisation de l'activité neuronale entre les individus diminue lorsque les individus ne parlent pas la même langue (Stephens et al., 2010). Autrement dit lorsqu'un individu écoute un partenaire, les aires dévolues à la perception de la parole s'activent de la même manière chez les deux individus - ce résultat étant certainement dû à la synchronisation des cortex auditifs sur le stimulus de parole - mais surtout, l'écoute de la parole provoque chez l'auditeur une activité oscillatoire qui permettrait d'anticiper la structure temporelle du message oral.

Les locuteurs se serviraient finalement de deux mécanismes distincts mais complémentaires pour se synchroniser et s'adapter dans l'interaction (Galantucci & Sebanz, 2009) : un mécanisme basé sur un phénomène d'imitation sensori-motrice exogène et endogène de type *bottom-up* et un mécanisme de synchronisation mutuelle des activités neuronales permettant la génération de prédictions temporelles de type *top-down*. Le modèle ADAM, que nous allons décrire dans la partie suivante, pourrait permettre d'expliquer

comment les partenaires parviennent à réajuster leur comportement pour faire face à la variabilité de leur partenaire et ainsi maintenir une bonne coordination.

3. Un modèle d'explication de l'adaptation dans l'interaction : ADAM

Comme nous l'avons décrit précédemment, lorsque deux partenaires entrent en interaction, ils exercent une influence réciproque sur leurs mouvements, leur parole et sur la dynamique de leur activité neuronale. Cette influence qualifiée de véritable attraction et contrainte par Fusaroli, Raczaszek-Leonardi et Tylén (2014), est un moyen pour les partenaires de réduire la variabilité intra-individuelle afin de se rendre plus prédictibles pour une meilleure coordination lors de l'interaction (Vesper, van der Wel, Knoblich, & Sebanz, 2012; Vesper et al., 2011). Néanmoins, une coordination réussie est non seulement basée sur la précision temporelle de l'anticipation des actions futures de son partenaire mais aussi et surtout sur l'adaptation mutuelle durant l'action actuelle. L'anticipation temporelle est un comportement prédictif lors duquel un partenaire va anticiper le déroulement temporel des actions ou de la parole de l'autre partenaire alors que l'adaptation est un comportement réactif lors duquel le partenaire va réajuster ses actions, sa parole, par un mécanisme de correction de l'erreur en fonction des variations temporelles de l'autre partenaire.

Les études menées sur la synchronisation sensori-motrice conjointe (i.e. tâche de tapping) ont pu montrer que l'accommodation temporelle à un partenaire se met en place très rapidement : les analyses de corrélation intertaps entre les partenaires montrent que ces derniers ont tendance à ajuster les intervalles temporels qui séparent deux de leurs taps pour s'approcher au mieux de ceux de leur partenaire dès les six premiers essais (corrélation positive à lag 0); et cet ajustement est bidirectionnel ce qui signifie qu'il n'y a pas un individu qui impose son tempo sans tenir compte de celui de l'autre (Konvalinka, Vuust, Roepstorff, & Frith, 2010). Cependant, lorsque le partenaire devient imprédictible (taps irréguliers) et qu'il ne s'adapte pas (i.e. ne tient pas compte des taps de l'autre individu), les capacités de synchronisation diminuent drastiquement.

Le modèle *ADaptation and Anticipation Model* (ADAM) de van der Steen et collaborateurs (2013) explique cette capacité à réajuster la programmation de ses propres

mouvements en fonction de ceux d'un partenaire en réunissant deux approches: celle de la correction réactive de l'erreur gérée par la phase des oscillateurs internes et basée sur le pourcentage d'asynchronie entre le tap présent et le tap précédent, et celle de l'anticipation gérée par l'horloge interne ou encore appelée timekeeper qui se base sur la différence entre l'Inter-Onset-Interval (IOI) actuel et l'intervalle précédent conservé par le timekeeper (van der Steen & Keller, 2013).

Dans ce modèle, la correction de phase est un phénomène d'ajustement local et automatique (inconscient) dépendant de l'activation des cortex auditifs et somato-sensoriels alors que la correction de période (qui intervient lorsque le changement est trop important) est dépendante de l'activation de l'aire motrice supplémentaire et nécessite une attention consciente sur les changements de tempo. Selon que l'on interagit avec un partenaire coopératif ou non, on va plutôt utiliser un mécanisme fixe de correction de phase ou une alternance entre correction de phase et de période. Repp et collaborateurs (2010) montrent que dans une tâche de synchronisation sensori-motrice, les réseaux neuronaux associés à l'adaptation temporelle et impliquant à la fois des processus attentionnel et anticipatoires seraient pour la correction de phase, l'activation des régions cérébelleuses liées aux aires motrices et auditives. Pour la correction de période, on retrouve l'activation des régions des ganglions de la base, du cortex préfrontal, des régions préfrontales, frontales médiales et pariétales (Repp, 2010). Les musiciens, et plus particulièrement les musiciens d'ensembles, sollicitent sans cesse ce type de mécanismes et de manière très précise étant donné à la fois la régularité temporelle et les variations de tempo imposées par la musique ce qui pourrait améliorer leurs compétences de coordination dans l'interaction. C'est ce que nous allons voir dans la partie suivante.

4. Les effets de la pratique du rythme musical sur les prédictions et l'adaptation temporelles en situation d'interaction

La pratique de la musique sollicite de manière spécifique et répétitive les mécanismes et propriétés neurophysiologiques permettant le développement de prédictions temporelles nécessaires au processus d'adaptation et à l'optimisation de la perception de la parole en situation d'interaction.

En effet, comme le montrent Novembre et Keller (2014) dans un article de revue, les musiciens accomplissent de manière répétitive des tâches qui requièrent la mise en correspondance de stimuli auditifs et de mouvements, soit une activation importante du système auditif et du système moteur ce qui a pour effet d'augmenter l'intégration sensori-motrice des stimuli (Novembre & Keller, 2014 ; voir aussi Karpati, Giacosa, Foster, Penhune, & Hyde, 2016). Le couplage renforcé entre ces aires permet aux musiciens de générer de meilleures prédictions de l'erreur lorsqu'ils produisent des mouvements. La pratique de la musique génère en effet chez eux des modèles internes plus efficaces sur les conséquences sensorielles de leurs gestes. Les tâches de synchronisation avec un métronome, qui demandent une coordination précise entre un stimulus auditif et un mouvement, ont montré que les musiciens sont plus consistants et plus précis que les non musiciens (Repp, 2010; Repp & Doggett, 2007). Mais les musiciens, possèdent également des représentations plus importantes de la structure des actions et peuvent alors anticiper avec plus de précision le but des actions des autres pour ensuite les intégrer à leurs propres actions (Novembre et Keller, 2014). Dans l'interaction, le couplage sensori-moteur est en effet un mécanisme essentiel qui permet à un individu d'intégrer les actions d'un autre individu (gestes, paroles) dans son propre répertoire moteur. Cette intégration permet de développer chez les deux individus en train d'interagir une représentation partagée de leurs actions et leur permet de parvenir à une compréhension sociale de de leurs actes (Knoblich & Sebanz, 2006). Autrement dit, une coordination inter-individuelle de bas niveau, basée sur des mécanismes sensori-moteurs, engendre une perception de plus haut niveau des actions du partenaire dans l'interaction. La coordination interindividuelle et la perception de haut niveau qui est d'ordre social, seraient améliorées par les mécanismes que sollicite la pratique de la musique (voir la notion de *Dyadic Motor Plan* développée par Sacheli, Arcangeli, & Paulesu, 2018).

En outre, lors des variations subites inhérentes à une situation d'interaction, les musiciens auraient également plus de facilités à s'adapter à des changements de tempo grâce à un meilleur entraînement de leurs oscillateurs internes sur les stimuli extérieurs mais aussi sur les rythmes endogènes de leurs partenaires. Loehr et Palmer (2011) ont demandé à des musiciens de reproduire des rythmes simultanément à l'écoute d'un métronome marquant la pulsation. Ce métronome a été manipulé afin de produire des variations de tempo au cours

de la reproduction rythmique. Les résultats sont mieux décrits par un modèle de type oscillateur (qui se base donc sur des relations de phase) que par un modèle de type *timekeeper* (qui se base sur les durées absolues) (Loehr, Large, & Palmer, 2011). Ceci suggère, comme le laissent entendre les modèles sur le tour de parole (Garrod & Pickering, 2015; Wilson & Wilson, 2005) que la coordination en musique serait basée sur l'entraînement d'oscillateurs internes sur des stimuli rythmiques externes. En outre, des études portant sur la production d'une œuvre musicale à deux musiciens ont montré que les musiciens qui possédaient les rythmes endogènes les plus proches, mesurés à partir de plusieurs tâches de tempo spontané, sont également ceux qui sont parvenus à se coordonner avec la plus grande précision durant la performance musicale conjointe (Zamm, Wellman, & Palmer, 2016). Ce résultat suggère que le couplage des activités neuronales inter-individuelles pourrait être un facteur influençant la coordination dans les actions conjointes telles que la musique d'ensemble ou le tour de parole.

En résumé, la réalisation d'une action conjointe telle que la conversation nécessite, comme la pratique de la musique d'ensemble, des capacités d'anticipation et d'accommodation qui sont rendues possibles par différents mécanismes développés par les individus en situation d'interaction : imitation exogène (i.e. convergence), imitation endogène (i.e. activation du réseau sensori-moteur), capacité d'entraînement sur un stimulus extérieur et synchronisation des activités neuronales inter-individuelles. Les musiciens, grâce à leur pratique répétitive et exigeante, sollicitent de manière plus importante ces mécanismes ce qui leur permet de se coordonner avec plus de précision lors de tâches rythmiques mais surtout de s'adapter avec plus de facilité lors des variations temporelles qui ont lieu dans les situations d'interactions.

III. Le développement normal de la parole, un processus qui passe par le rythme de l'interaction

A la lumière des deux chapitres précédents, nous pouvons affirmer que pouvoir se synchroniser sur l'enveloppe temporelle de la parole et structurer les événements auditifs qui se déroulent dans le temps sont des processus essentiels pour la perception de la parole mais également pour la coordination avec un interlocuteur lors d'une conversation.

L'analyse des premiers échanges kinesthésiques et vocaux entre l'adulte et l'enfant, semblent montrer que ces processus se mettent en place très précocement et que ce "cadre" temporel est important pour l'enfant puisse développer la compréhension des unités de parole (Nazzi & Ramus, 2003) et devenir un partenaire conversationnel adapté.

1. Développement précoce de la perception et de la production de structures temporelles

La sensibilisation de l'être humain à la rythmicité commence dès l'exposition du fœtus à ses propres activités physiologiques primaires : battements cardiaques, succion, respiration, hoquet, déglutition mais également par la perception des rythmes physiologiques de sa mère. Grâce à la mesure des variations de son rythme cardiaque, il est en effet possible de mettre en évidence que dès 32 semaines d'âge conceptuel (AC), le fœtus réagit aux perturbations du rythme cardiaque de sa mère (Monk et al., 2000). La mesure des changements du rythme cardiaque du fœtus nous montre également que ce dernier est sensible à des variations rythmiques provenant de la perception de voix telles qu'elles lui parviennent par voie externe (Provasi, Anderson, & Barbu-Roth, 2014). La perception des indices rythmiques de nature verbale va continuer à s'affiner pour permettre au fœtus, entre 32 et 34 semaines d'AC, de distinguer sur la base de leurs caractéristiques prosodiques d'origine temporelle, la voix de sa mère de la voix d'une autre femme dont la fréquence fondamentale est par conséquent voisine de celle de sa mère (Kisilevsky et al., 2009).

Cette sensibilité précoce du fœtus aux caractéristiques temporelles du langage, va se refléter à la naissance. En effet, en enlevant les informations acoustiques correspondant aux informations phonétiques contenues dans le signal de parole, des études ont pu mettre en

évidence que non seulement les nouveau-nés préfèrent écouter des patterns rythmiques accentuels qui proviennent de leur langue maternelle (Mehler et al., 1988) mais parviennent également, à quelques jours de vie seulement, à discriminer des langues à partir de leurs caractéristiques rythmiques (Nazzi, Bertoncini, & Mehler, 1998). Chez les bébés, l'intégration de patterns réguliers dans la parole est mise en évidence par la production de cris dont les contours mélodiques vont varier en fonction de la langue maternelle à laquelle ils appartiennent. Ainsi, les bébés allemands vont produire des cris avec des patterns mélodiques différents de ceux des bébés français. Alors que les premiers produisent un contour mélodique associant un ton montant puis un ton descendant, typique de celui le plus fréquemment retrouvé dans la prosodie de l'allemand chez l'adulte, les seconds produisent un contour mélodique associant un ton descendant puis un ton montant, retrouvé sur les frontières de constituants du français (excepté celui de frontière majeure qui est descendant) (Mampe, Friederici, Christophe, & Wermke, 2009). On sait également que les bébés de 9 mois préfèrent écouter des extraits de parole dont les pauses préservent un pattern métrique connu dans leur langue (dans cette étude pattern trochaïque) plutôt que des stimuli verbaux dont les pauses rompent ce pattern rythmique (Echols, Crowhurst, Childers, Becker, & Rader, 1997). Des bébés âgés de 8,5 mois préfèrent également écouter des extraits de parole avec des pauses qui correspondent à des frontières de phrases plutôt que des pauses à l'intérieur des phrases (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). La propension des bébés en période prélinguistique à grouper des unités de paroles en patterns, suggère que ce processus rythmique est sûrement déjà présent in utero.

Ces résultats vont de pair avec la capacité des enfants d'âge préverbal à percevoir des unités régulières i.e. des *beats* dans la parole. Dans les années 70, grâce à l'analyse de vidéos de nouveau-nés couplés à l'analyse d'oscillogrammes de la parole, des chercheurs ont pu mettre en évidence que lorsque des nouveau-nés en mouvement perçoivent la parole d'un adulte, ils produisent des changements dans leurs micro-mouvements corporels (i.e. mouvements des sourcils, des épaules, des pieds, des hanches) au moment de l'occurrence des phonèmes et des syllabes de l'adulte ; cet alignement des mouvements sur la parole est observable que l'adulte soit présent ou bien en réaction à de la parole enregistrée (Condon & Sander, 1974). On sait de plus que les enfants âgés de 11 à 24 semaines, peuvent discriminer

si des clicks sont alignés avec les P-centers de la parole plutôt qu'avec l'onset des syllabes. Fowler a pu montrer dans une expérience complémentaire que le pattern de résultats observé était congruent avec le fait que les enfants perçoivent comme isochrones des syllabes alignées sur le p-center et anisochrones des syllabes alignées sur l'onset. En effet, le temps d'habituation plus long est observé pour des clicks alignés sur le p-center et de la parole isochrone à partir des P-centers. Dans l'expérience contrôle le temps d'habituation le plus long est observé pour des clicks alignés sur le p-center et de la parole objectivement isochrone, car composée d'une même syllabe (Fowler, Smith, & Tassinary, 1986). La synchronisation spontanée des mouvements sur la parole ainsi que la perception des P-centers comme beat rythmique dans la parole suggèrent que, comme le font les adultes, les enfants encore à un stade préverbal, sont capables d'extraire une pulsation régulière dans la parole en se basant sur des points d'ancrage perceptifs (cf. 1^{ère} partie page 3).

La sensibilité précoce des nouveaux-nés à la temporalité présente dans la parole, va se manifester de manière encore plus prégnante lors de l'écoute d'autres types de stimuli réguliers et en particulier lors de l'écoute de la musique. Comme chez les adultes, le rythme musical, de par sa régularité intrinsèque et la répétition de cette régularité, va favoriser chez les bébés la synchronisation des mouvements sur le stimulus acoustique et l'émergence de structures temporelles. Ainsi, les enfants âgés entre 5 et 24 mois produisent plus de mouvements rythmiques lors de l'écoute de la musique ou d'un métronome comparé à l'écoute de la parole (Zentner & Eerola, 2010). Même si les capacités de synchronisation sont très dépendantes de la maturation du système sensori-moteur et que la précision de la synchronisation sur du matériel musical ne serait présente qu'à partir de 4 ans (McAuley, Jones, Holub, Johnston, & Miller, 2006), une récente étude (Fujii et al., 2014) a pu mettre en évidence, à l'aide de capteurs de mouvements, que les bébés, dès 3-4 mois, synchronisent leurs mouvements de jambes et leurs vocalisations lorsqu'ils entendent de la musique. Les bébés seraient donc capables, très précocement, d'extraire une pulsation dans la musique comme dans la parole, mais ils sont aussi capables de détecter et d'adapter leurs mouvements à des changements de tempo liés aux variations de la pulsation (Zentner & Eerola, 2010 ; Baruch & Drake, 1997). Des mesures électrophysiologiques confirment la

capacité précoce des bébés à extraire la pulsation. Winkler et collaborateurs (2009) ont montré grâce à une mesure de négativité de discordance (Mismatch Negativity: MMN) que les nouveau-nés sont capables de détecter l'omission de la pulsation dans des séquences sonores musicales et que ces bébés, âgés de seulement quelques jours, réagissent plus particulièrement à l'omission du *downbeat* qui est la pulsation la plus importante dans la structure hiérarchique des structures rythmiques musicales. Ce dernier résultat suggère que les bébés ont été sensibles à l'organisation des pulsations selon une structure hiérarchique. Les bébés seraient en effet sensibles à différents niveaux de hiérarchie dans le stimulus acoustique. Ils préfèrent ainsi écouter des séquences de notes lorsque celles-ci sont organisées selon une structure métrique et particulièrement lorsque cette structure métrique est forte comparé à leur présentation en l'absence de structure ou d'une structure métrique plus faible (Bergeson & Trehub, 2006; Hannon & Trehub, 2005).

L'induction d'une pulsation dans la musique ainsi que la perception auditive précoce de patterns temporels musicaux sont des compétences temporelles présentes précocement chez l'enfant. Leur développement semble néanmoins facilité par la mise en mouvement. En effet, lorsque des enfants 7 mois entendent une suite de six pulsations non accentuées puis qu'ils sont habitués à être bercés, pendant deux minutes, toutes les 2 ou toutes les 3 pulsations, les enfants bercés toutes les 3 pulsations préféreront ensuite écouter la séquence initiale de six pulsations, accentuée selon une métrique ternaire plutôt que binaire (Phillips-Silver & Trainor, 2005). Ces résultats vont dans le sens d'une récente étude (Cirelli, Spinelli, Nozaradan, & Trainor, 2016) montrant, à l'aide de mesures électrophysiologiques (Steady-State Evoked Potentials : SSEP), que les bébés âgés de 7 mois présentent, comme les adultes, des réponses neuronales qui reflètent la fréquence d'apparition de la pulsation et de la métrique lors de l'écoute de séquences isochrones de tons non accentués, donc ne marquant pas acoustiquement la métrique dans le stimulus (Cirelli et al., 2016; voir Nozaradan, Peretz, Missal, & Mouraux, 2011 pour l'adulte) (voir figure 3.1). Ces résultats suggèrent que les bébés comme les adultes développent précocement une cognition de type rythmique qui leur permet de percevoir des structures temporelles dans les stimuli acoustiques.

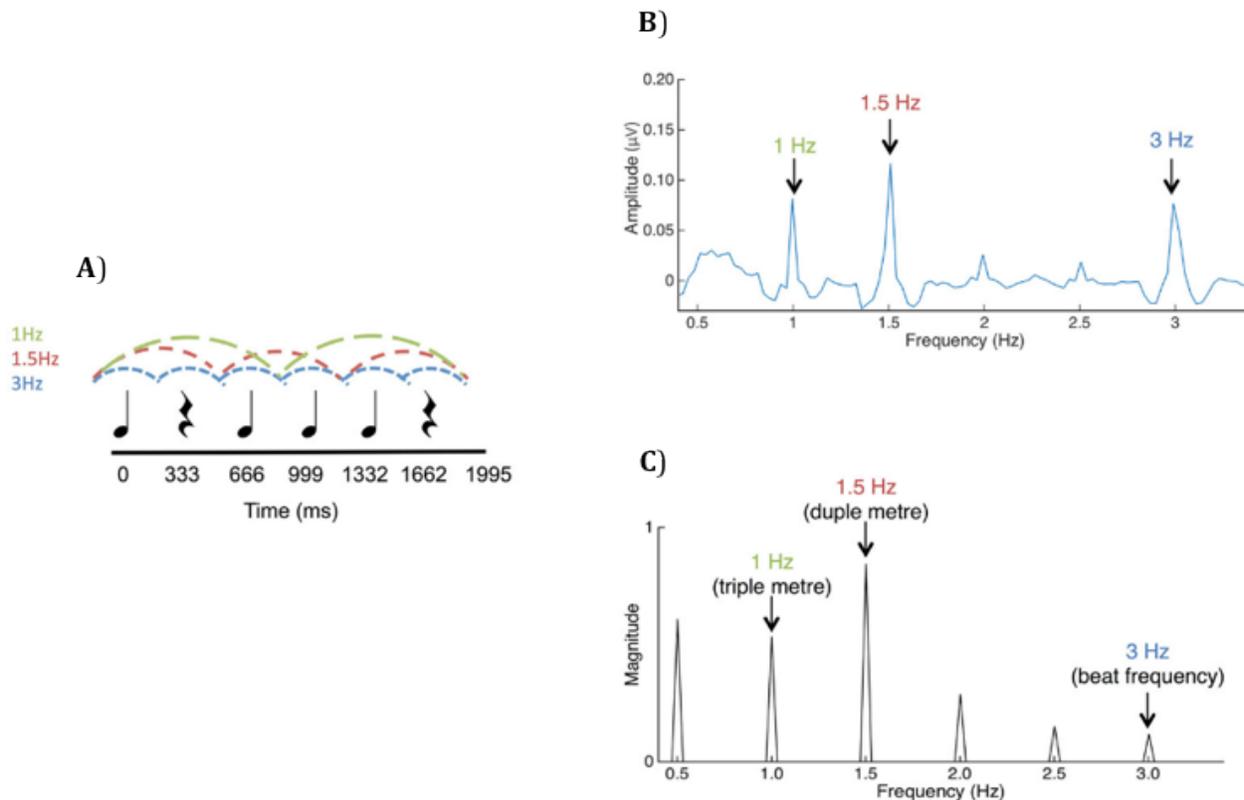


Figure 3.1. Représentation des réponses neuronales des enfants mimant la structure rythmique du stimulus au niveau du beat et du mètre extraite de Cirelli et collaborateurs (2016). A) Notation musicale du stimulus auditif composé d'une succession de tons et de silences espacés de manière isochrone. Sont représentés en bleu, le beat (3Hz), en rouge la structure métrique binaire (1.5 Hz) et en vert la structure métrique ternaire (1Hz). B) Spectre de l'enveloppe temporelle C) Moyenne des réponses évoquées de tous les enfants dans laquelle on retrouve les fréquences du beat et des structures métriques binaire et ternaire.

Dans la partie suivante, nous allons voir que la sensibilité du bébé aux structures temporelles dans la parole et dans la musique, vraisemblablement déjà en développement in utero, va progressivement s'étendre au cadre de la conversation grâce aux interactions précoces entre l'adulte et l'enfant.

2. De l'émergence de la rythmicité et des prédictions dans l'interaction

En plus d'être sensible à la temporalité dans la musique et la parole, le bébé va développer une sensibilité accrue à la temporalité de l'interaction.

Quand un bébé émet un signal de communication, il s'attend à ce que l'adulte lui réponde mais dans un laps de temps bien spécifique : entre 1 et 2 secondes (voir Keller, Lohaus, Völker, Cappenberg, & Chasiotis, 1999). Cette capacité de l'adulte à répondre dans un temps imparti aux initiatives de l'enfant est appelée *contingence temporelle*. Si ce temps de réponse n'est pas respecté, le bébé semble ne pas considérer les comportements de l'adulte comme inscrits dans un espace intersubjectif commun ; autrement dit, il ne ressent pas que l'adulte lui prête des intentions de communication et risque de ne pas développer ses compétences langagières de manière optimale. La contingence temporelle des réponses de l'adulte conditionne en effet le développement des vocalisations du bébé (Kuhl, 2012) ainsi que celui des comportements linguistiques de plus en plus complexes du jeune enfant (M. H. Goldstein, King, & West, 2003; Michael H. Goldstein & Schwade, 2008). Lorsque des bébés âgés de 6 à 12 semaines interagissent en temps réel avec leur mère à travers une télévision, puis que l'on insère un délai de seulement une seconde dans la transmission de l'image de la mère, les bébés se désintéressent de l'interaction. En effet, du fait de l'interaction directe, les mères conservent bien une attitude souriante et bienveillante et adressée au même enfant mais ne répondent pas, à cause du délai de transmission de l'image, de manière contingente aux actions et productions du bébé, ce qui engendre chez ce dernier un arrêt de l'interaction ici manifesté par une réduction du taux des regards vers la vidéo (Henning & Striano, 2011; Nadel, Carchon, Kervella, Marcelli, & Reserbat-Plantey, 1999). Les bébés âgés de 3 et 4 mois, produisent en outre des vocalisations plus similaires à des sons de la parole (versus des bruits de bouche) quand les réponses de l'adulte sont temporellement contingentes à leurs productions. De plus, dans cette condition, leurs vocalisations sont séparées par de longs intervalles. Le délai plus important laissé par le bébé entre deux vocalisations suggère, que dans la situation où l'adulte fournit des réponses temporellement contingentes, il considère que l'adulte a l'intention d'échanger avec lui et s'attend, dans ce cadre-là seulement, à une probable réponse de sa part (Masataka, 1993).

La sensibilité temporelle et les prédictions précoces développées par le bébé dans le cadre interactionnel pourraient expliquer comment les enfants parviennent à se coordonner avec un adulte qui interagit avec eux. Autour de deux mois, le bébé est en effet capable de produire des vocalisations en alternance avec la parole de l'adulte en marquant des pauses régulières de 500 ms à 1 seconde entre les échanges (Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001). Une récente étude (Dominguez, Devouche, Apter, & Gratier, 2016) a révélé que dès 2 à 4 jours de vie, les nouveau-nés sont eux-aussi capables de s'engager dans des échanges de vocalisations avec leur mère, marqués par une véritable alternance. Ainsi, ces bébés vocalisent la plupart du temps dans la seconde qui suit l'arrêt des vocalisations de la mère et lui répondent même souvent dans les 50 ms qui suivent la fin de son tour de parole. Ce résultat suggère que les nouveau-nés sont déjà capables d'anticiper la fin du tour de leur mère avec précision. En outre, l'analyse de l'organisation temporelle de ces échanges préverbaux, montre que lorsque la mère est contingente, le bébé s'implique activement dans les échanges qui prennent alors l'allure de véritables tours de parole.

Ces échanges précoces entre l'adulte et l'enfant, même s'ils ne sont pas encore constitués de productions verbales sont suffisamment structurés temporellement pour que l'on perçoive en effet l'émergence de ce qui va devenir le tour de parole. Néanmoins, comme nous l'avons vu au chapitre 2, le tour de parole est une activité qui demande d'importantes capacités d'anticipation et d'adaptation liées en partie à la maturation du système sensori-moteur. La capacité à se coordonner dans le tour de parole va donc s'affiner progressivement grâce aux nombreux ajustements que la mère va mettre en place au cours des échanges précoces et aux différents partenaires conversationnels que va rencontrer l'enfant (Casillas, 2014). Ainsi, chez des enfants âgés de 7 à 18 mois, la proportion de chevauchements (i.e. les vocalisations de l'enfant produites en même temps que la parole de l'adulte) tend à diminuer avec l'âge, en particulier vers 12 mois, âge auquel les vocalisations seront produites en alternance dans des proportions plus importantes (Ginsburg & Kilbourne, 1988). Une autre étude longitudinale, menée sur des enfants âgés de 3 à 18 mois en interaction avec leur mère a montré que le pourcentage de chevauchements diminue avec l'âge mais plus précisément que ce pourcentage se réduit presque de moitié entre 3 et 9 mois (Gratier et al., 2015) pour devenir identique à celui de la mère à l'âge de 18 mois. Ce résultat, obtenu avant

9 mois, suggère que la réduction du pourcentage de chevauchements ne peut pas être attribuable à une augmentation de la compréhension des paroles de l'adulte par l'enfant mais plutôt à la capacité de l'enfant à s'inscrire dans une véritable temporalité de l'alternance dans l'interaction. En outre, un autre résultat de cette étude montre que lorsque les enfants alternent leurs vocalisations avec celles de l'adultes, la durée des silences entre les tours est plus courte à 5 mois comparée à 9 et 12 mois ce qui pourrait signifier qu'avant 9 mois, les enfants répondent à l'adulte dans un "simple" jeu d'alternance temporelle de donner-recevoir, mais qu'à partir 9 mois, âge de l'apparition de l'*attention conjointe*, entre en jeu le développement de compétences sémantiques qui ralentissent leur temps de réponse et augmente la durée des silences entre les tours. L'échange semble, à partir de cette période, être considéré par l'enfant comme un espace pour partager des informations sur le monde environnant, les objets et sur les émotions.

Le bébé possède donc bien, dès quelques jours de vie, des capacités de coordination temporelles avec l'adulte lors d'une interaction. Même si au cours du développement, l'alternance vocale bébé/adulte est soumise à des variations temporelles dues à l'émergence de capacités linguistiques, il n'en reste pas moins que cette alternance n'est possible que si le bébé est capable d'anticiper une réponse de la part de l'adulte et plus précisément la fin de son tour de parole. Pour permettre à l'enfant d'anticiper au mieux la réponse de l'adulte d'un point de vue temporel, ce dernier va systématiquement réajuster son comportement verbal ou non verbal en fonction des réponses de son enfant, afin que l'échange conserve un certain degré de rythmicité. En effet, contrairement à l'enfant, la mère va être très "permissive" quant aux variations de contingences temporelles de l'enfant (Henning & Striano, 2011; Striano & Stahl, 2006) dont les réponses peuvent être jusqu'à 10 fois plus longues chez les enfants de 3 ans et 5 ans par rapport aux moyennes des latences de réponses des adultes (Berninger & Garvey, 1981; Lieberman & Garvey, 1977 cité dans Casillas et al., 2015). Autrement dit, la mère va poursuivre l'échange même si son enfant ne produit pas une réponse contingente et va se réajuster afin que l'échange conserve une régularité permettant le développement de prédictions temporelles chez l'enfant.

Des études utilisant la technique de la poursuite des mouvements oculaires (eye-tracking) ont pu mettre en évidence les capacités d'anticipation des jeunes enfants dans le

tour de parole. Le paradigme expérimental de ces études consiste à équiper les enfants d'un système d'eye-tracking et de leur présenter un dialogue entre deux personnages ou des marionnettes. La mesure des mouvements oculaires permet d'analyser à quel endroit l'enfant porte son regard au cours du dialogue et plus particulièrement au moment où les interlocuteurs vont changer de tour de parole. Si l'enfant regarde le locuteur qui va prendre son tour de parole alors que le locuteur précédent n'a pas encore terminé son énoncé mais s'apprête à le faire, c'est probablement qu'il a anticipé la fin du tour de parole du locuteur en cours (voir figure 3.2). Une première étude réalisée par Keitel et collaborateurs (2013) a montré que parmi quatre groupes d'enfants âgés de 6-12-24 et 36 mois, seuls ceux de 36 mois étaient capables d'orienter leur regard par anticipation sur le locuteur à venir (Keitel, Prinz, Friederici, Hofsten, & Daum, 2013). Une étude de Casillas et Franck (2017) a montré, à l'aide du même type de paradigme que les enfants sont capables d'anticipation à partir de deux ans et que cet effet plus marqué sur les paires de questions-réponses, augmente avec l'âge. Autrement dit, dès 2 ans, les enfants seraient capables de prédire que le locuteur en cours va terminer son tour de parole et cette anticipation serait facilitée lorsque le locuteur en cours pose une question qui appelle une réponse de la part de l'interlocuteur ; ce type de paires d'actes de parole est en effet hautement prédictible (Casillas & Frank, 2017). Contrairement aux adultes qui se basent essentiellement sur des indices lexico-syntaxiques (De Ruiter et al., 2006; ; voir aussi Keitel & Daum, 2015 pour un résultat différent), les enfants s'appuieraient, en fonction de leur âge développemental, sur une association d'indices lexico-syntaxiques et prosodiques pour anticiper la fin du tour de parole (Casillas & Frank, 2017; Keitel & Daum, 2015).



Figure 3.2. Représentation de la mesure de l'anticipation temporelle des enfants, lors de l'observation de dialogues adaptée de Casillas et Franck (2017). La marionnette n°1 parle pendant que la marionnette n°2 écoute. Dans le cas d'une anticipation du tour de parole, l'enfant, équipé d'un eye-tracker, va déplacer et fixer son regard sur la marionnette n°2 alors que la marionnette n°1 n'a pas encore terminé de parler ; c'est ce que représentent les points de couleur bleue, violette et rouge sur la marionnette n°2.

En résumé, les jeux de tours de parole/rôle avec l'adulte dans lesquels le bébé s'inscrit sont au départ maintenus sans une volonté de partager des informations avec l'adulte, mais ces jeux de donner-recevoir sont néanmoins temporellement structurés et établissent le cadre dans lequel vont émerger chez l'enfant des acquisitions linguistiques de plus en plus complexes. Autrement dit, c'est dans ce cadre interactionnel où se construit une synchronie interactionnelle, que l'enfant et l'adulte vont pouvoir partager des représentations temporelles et conceptuelles communes.

IV. Les capacités de prédictions des enfants sourds, une pièce manquante dans l'explication des déficits perceptifs des enfants sourds

Les enfants sourds congénitaux, implantés cochléaires, sont une population vulnérable en termes de perception et production du langage, et plus particulièrement en situation d'interaction. En effet, même si l'implant cochléaire améliore considérablement la perception de la parole, les niveaux de langage atteints par ces enfants demeurent très hétérogènes et certaines situations de communication détériorent les résultats obtenus en cabine audiométrique ou en face à face au calme, avec un seul interlocuteur. Dans ce chapitre, nous allons décrire le fonctionnement technique de l'implant cochléaire (IC) ainsi que les limitations perceptives dues à cette technologie. Nous décrirons ensuite l'évolution et les difficultés langagières des enfants implantés cochléaires ainsi que les possibles causes de l'hétérogénéité des niveaux de langage retrouvés post-implantation. Enfin, nous relaterons les quelques études qui ont analysé les effets d'une stimulation musicale sur les capacités de perception et de production de la parole chez les enfants implantés cochléaires.

1. Mise en place, fonctionnement et perception auditive avec un l'implant cochléaire

La surdité congénitale touche environ une naissance sur 1000. Parmi ces surdités, la surdité profonde touche 84 % des enfants sourds. Suite au dépistage qui a lieu en maternité à 3 jours de vie puis au diagnostic réalisé à l'hôpital ou en cabinet libéral en moyenne entre 3 et 6 mois de vie, les enfants dont les familles en font le choix, sont équipés de prothèses auditives conventionnelles pendant une durée de 4 à 6 mois. Cet appareillage par voie aérienne stéréophonique qui sera posé en rétro auriculaire si le bébé tient sa tête ou en Y si ses pavillons sont mous, permet de retransmettre sans les déformer et en les amplifiant, les signaux acoustiques émis dans l'environnement. A la fin de ces 6 mois de port de la prothèse conventionnelle et suite à un bilan réalisé par une équipe pluridisciplinaire incluant orthophonistes, audioprothésistes, psychologues et médecin ORL, une implantation cochléaire et une prise en charge peuvent être proposées à la famille. L'implantation est

acceptée par la famille, elle sera réalisée en moyenne avant l'âge de 2 ans. La plupart des enfants atteints de surdité profonde et sévère seront opérés sur une seule oreille et porteront un implant avec ou sans prothèse controlatérale sur l'oreille non opérée. Dans des cas de cophose bilatérale, de surdité évolutive ou de risque d'ossification de la cochlée suite à une méningite, les enfants se verront équipés de deux implants posés de manière simultanée mais le plus souvent, cette double implantation se fera de manière séquentielle. L'implant cochléaire (figure 4.1) à la différence de la prothèse conventionnelle, nécessite une opération chirurgicale lors de laquelle le chirurgien insère à l'intérieur de l'organe sensoriel de l'oreille, la cochlée, une rangée d'électrodes (4) ; le nombre d'électrodes varie en général de 12 à 22 électrodes. Ces électrodes vont avoir pour fonction de transmettre des impulsions électriques au nerf auditif (en jaune sur la figure) qui enverra ensuite ces informations au cerveau. Les impulsions électriques proviennent d'un processeur (1) posé sur l'oreille de l'enfant qui transforme les sons de l'environnement en signal électrique en privilégiant les fréquences émises par la parole qui s'étalent entre 125 et 8000 Hz.

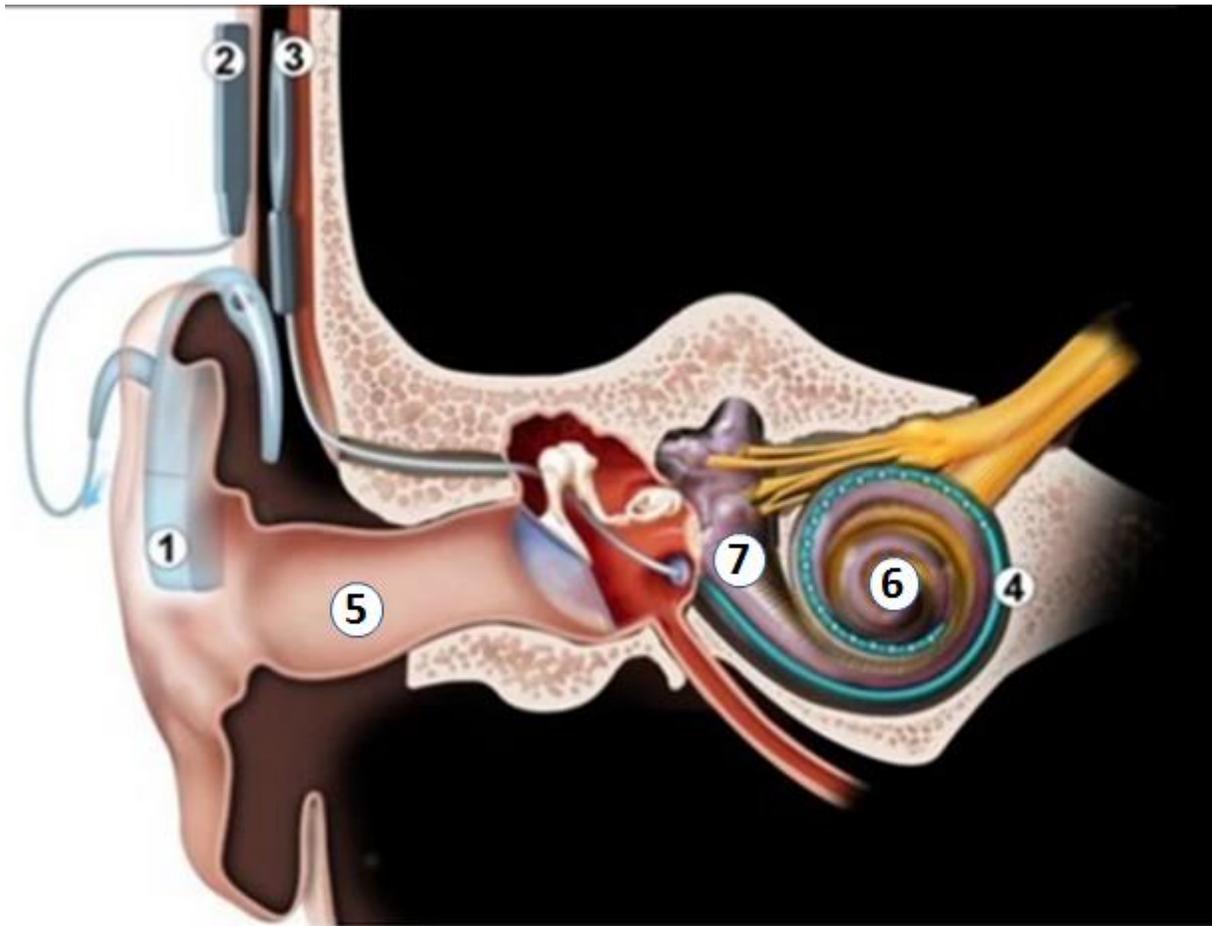


Figure 4.1. Représentation de l'oreille et de l'implant cochléaire extraite du site voyage au centre de l'audition. <http://www.cochlea.eu/rehabilitation/implants-cochleaires>

Le signal, envoyé au nerf auditif via l'implant cochléaire est bien différent de celui envoyé par la prothèse conventionnelle. L'implant cochléaire, contrairement à la prothèse "se substitue" aux fonctions de l'oreille dont le rôle est de transformer les informations acoustiques qui entrent dans le conduit auditif externe (5) en informations mécano-électriques ; cette transformation réalisée par une oreille saine permet entre autres de fournir des informations redondantes pour le codage et une perception précise des fréquences. L'oreille interne utilise en effet deux stratégies pour coder les fréquences contenues dans les stimuli acoustiques : le codage spatial dit "de place", lors duquel les sons parvenus au tympan vont activer, via les mouvements d'une membrane contenue dans la cochlée (membrane basilaire), environ 3500 cellules sensorielles (pour les cellules ciliées

internes) réparties tout le long de la cochlée. Ces cellules vont coder différentes fréquences (de 20 à 20000 Hz) en fonction de leur emplacement le long de la cochlée. Ainsi, les sons avec des fréquences graves seront codés par les cellules situées à l'apex de la cochlée (6) alors que les sons aigus seront codés par les cellules situées à sa base (7). Cette répartition spatiale des cellules ciliées en fonction des fréquences est nommée tonotopie et se retrouve également tout le long des voies auditives jusqu'au cortex auditif primaire (aire 41 de Brodmann). L'autre stratégie utilisée par l'oreille pour le codage des fréquences est la stratégie temporelle dite "principe de la volée", lors de laquelle les neurones du nerf auditif se synchronisent sur la période ou sur des multiples entiers de la période du stimulus acoustique. Cette autre stratégie a cependant des limites physiologiques qui ne permettent pas de percevoir la fréquence au-delà de 5000 Hz. Ces deux types de stratégies se complètent pour permettre une perception précise de la hauteur des sons qui permet de discriminer les différents tons dans la musique ou les subtiles variations prosodiques dans la parole. Elles permettent également de traiter la structure dite "fine" de la parole grâce à laquelle nous pouvons discriminer des phonèmes qui diffèrent en termes de lieu d'articulation ou de voisement (cf. partie n°1 pages 13) ou de percevoir la parole dans le bruit.

Du fait de ses caractéristiques techniques, l'implant cochléaire limite beaucoup l'utilisation de ces deux stratégies du codage des fréquences. En effet, l'implant va séparer le signal acoustique en différentes bandes de fréquences puis extraire l'enveloppe temporelle du signal obtenu. Autrement dit, l'implant va extraire les variations d'amplitude lentes du signal acoustique (2 à 50 Hz) mais pas la structure fine. Après une étape de compression, l'enveloppe temporelle sera transmise aux électrodes qui enverront les impulsions au nerf auditif. Ces électrodes, bien que respectant la tonotopie de la cochlée ne sont qu'au nombre de 22 (nombre maximum contre 3500 cellules ciliées internes) et ne vont donc permettre qu'une transmission imprécise des fréquences contenues dans le signal acoustique.

L'implant cochléaire réduit également le codage des variations d'intensité contenues dans le signal. Ainsi, s'il est possible de discriminer de 60 à 100 paliers différents sur une échelle de 0 à 120 dB pour une oreille saine l'implant cochléaire ne permet de discriminer seulement 20 paliers pour une échelle de 6 à 30 dB (pour plus de détails voir Limb & Roy, 2014).

Même si l'implant cochléaire favorise le codage des bandes de fréquences contenues dans la parole, il ne permet pas de percevoir le timbre vocal qui dépend d'une association du codage des informations fréquentielles, temporelles et d'intensité (codage spectral) mais il limite également la perception des informations prosodiques et des informations phonétiques relevant d'indices temporels rapides. Ces limitations vont avoir des conséquences sur la perception du langage et particulièrement dans les situations d'écoute nécessitant de séparer plusieurs flux acoustiques produits simultanément (cocktail party).

Ces limitations de codage de l'information acoustique et en particulier de la structure fine vont en outre dégrader l'appréciation de la musique par les personnes implantées cochléaires, en tout cas pour celles qui sont devenues sourdes et qui ont eu un passé auditif musical sans implant cochléaire. L'implant va limiter le seuil de discrimination des tons ($\frac{1}{2}$ ton pour une personne normo-entendante contre 1 à 8 demi-tons pour les personnes IC) et ainsi engendrer chez les personnes implantées cochléaires des difficultés à identifier les variations mélodiques de la musique (combinaison séquentielle des tons) surtout lorsque les tons présentés sont des tons complexes comparé à des tons purs (Galvin, Fu, Shannon, & Shannon, 2009). L'IC va aussi limiter l'appréciation de l'harmonie (combinaison simultanée des sons). La perception du timbre musical va également être altérée et engendrer chez la personne implantée cochléaire des difficultés à discriminer différents instruments. Cependant, les paramètres technologiques de l'implant ne semblent pas limiter la perception des informations temporelles musicales telles que le rythme. Les adultes devenus sourds implantés cochléaires se basent d'ailleurs sur le rythme musical pour reconnaître ou discriminer des mélodies (Kong, Cruz, Jones, & Zeng, 2004). On sait en outre que les enfants sourds congénitaux implantés cochléaires fournissent une reproduction plus fidèle des caractéristiques temporelles des chansons comparé à leurs caractéristiques fréquentielles lorsqu'ils chantent des chansons connues (Nakata, Trehub, Mitani, & Kanda, 2006).

2. Conséquences de la surdité et de l'implantation cochléaire sur le développement du langage

Bien que dans l'ensemble, l'implantation précoce dans le cadre d'une surdité congénitale permette aux enfants sourds profonds d'intégrer des écoles ordinaires et de vivre dans un milieu entendant, les résultats des études montrent que leurs niveaux de langage sont assez hétérogènes. Malgré de nombreuses recherches sur la source de cette variabilité, il demeure difficile de déterminer quel (s) est (sont) le (s) prédicteur (s) d'un bon niveau de langage post-implantation (van Wieringen & Wouters, 2015).

On sait que l'implantation, en particulier l'implantation avant l'âge de 2 ans permet aux enfants sourds profonds d'obtenir des niveaux de langage plus proches de ceux de leurs pairs normo-entendants comparé aux résultats obtenus à partir de prothèses conventionnelles (Geers, 1997; Truy et al., 1998). Parmi les enfants implantés cochléaires, l'âge d'implantation, les stratégies de codage de l'information acoustique par l'implant, le mode de communication, mais également le respect par les familles des étapes chronologiques menant du dépistage à la prise en charge précoce (1-3 et 6 mois aux Etats-Unis) sont évoqués comme de possibles prédicteurs des scores de langage post-implantation (Dowell, Dettman, Blamey, Barker, & Clark, 2002; Yoshinaga-Itano, Sedey, Wiggin, & Chung, 2017). L'âge d'implantation et la durée du port de l'IC semblent être de bons prédicteurs avec un effet particulièrement important sur les scores de langage en perception et production (Dettman et al., 2016). Niparko et collaborateurs (2010) ont évalué le développement linguistique sur 3 ans d'une grande cohorte d'enfants implantés avant 5 ans (N= 188) et recrutés dans 6 centres d'implantation différents. Des mesures langagières en expression et compréhension, des mesures de reconnaissance de la parole mais aussi de la quantité des interactions parent-enfant ont été collectées à 6-12-24 et 36 mois post-implantation. Ces mêmes mesures ont été relevées chez un groupe contrôle d'enfants normo-entendants. Les résultats confirment l'effet d'une implantation précoce et de la durée de la déprivation sensorielle sur la rapidité du développement du langage. L'audition résiduelle, le taux d'interactions parent-enfant et le niveau socio-économique des familles sont également des facteurs associés à l'amélioration des capacités linguistiques au cours du temps (Niparko et al., 2010).

L'implant cochléaire peut ainsi permettre aux enfants sourds de considérablement améliorer leur niveau de langage en 3 ans, mais il ne leur permet pas toujours de rattraper le niveau de leurs pairs normo-entendants (Boons et al., 2013; Geers, Nicholas, & Sedey, 2003; Niparko et al., 2010). Même équipés d'un implant cochléaire bilatéral, les enfants obtiennent des résultats très variables en termes de niveaux de langage et ces différences inter-individuelles ne seraient pas expliquées par la durée du port de l'un ou des deux implants, ni par le niveau d'éducation des familles (Hess, Zettler-Greeley, Godar, Ellis-Weismer, & Litovsky, 2014).

Plus précisément, comparé aux scores de leurs pairs normo-entendants, les enfants sourds IC présentent des déficits en fluence verbale phonologique et sémantique (Wechsler-Kashi, Schwartz, & Cleary, 2014) ainsi qu'en compréhension lexicale. Un niveau de langage plus particulièrement affecté est le niveau de production et de compréhension morphosyntaxique ; ces difficultés semblent liées à des déficits d'attention auditive. En effet, les scores de compréhension morphosyntaxique sont encore plus faibles lorsque les tâches nécessitent la discrimination de mots proches d'un point de vue phonologique (Caselli, Rinaldi, Varuzza, Giuliani, & Burdo, 2012). Les déficits en compétences syntaxiques semblent avoir un impact négatif sur les compétences métacognitives des enfants implantés telles que l'accès à théorie de l'esprit ou encore sur le développement de la compréhension en lecture (Lederberg et al., 2013). Des compétences linguistiques plus élaborées telles que la compétence narrative (Boons et al., 2013), discursive ou le raisonnement verbal abstrait sont des domaines où l'enfant sourd, même implanté cochléaire, rencontre des difficultés comparé à ses pairs normo-entendants (Geers, Moog, Biedenstein, Brenner, & Hayes, 2009; Geers & Sedey, 2011). La prosodie est un autre niveau de langage particulièrement déficitaire chez l'enfant implanté cochléaire. Comme expliqué dans la première partie de ce chapitre, le codage de l'information acoustique par l'implant cochléaire réduit drastiquement le nombre et la précision des fréquences perçues, ce qui se traduit par des limitations dans la perception fine des indices prosodiques. Les enfants implantés cochléaires, comparé à leurs pairs normo-entendants, présentent en effet plus de difficultés dans les tâches de discrimination d'intensité, de durée et de fréquence fondamentale. Ces difficultés sont associées à des scores plus faibles que le leurs pairs normo-entendants dans des tâches de discrimination de mots

et de phrases en fonction de l'accent (Torppa, Faulkner, et al., 2014). Par ailleurs, la capacité à discriminer des énoncés sur la base des émotions contenues dans la voix est faible voire impossible (pour la colère versus la joie et la tristesse) chez des enfants implantés âgés entre 5 et 13 ans comparé à un groupe d'enfants normo-entendants (Nakata, Trehub, & Kanda, 2012 ; voir également Jiam, Caldwell, Deroche, Chatterjee, & Limb, 2017). Les adolescents implantés cochléaires utilisent aussi moins efficacement que les adolescents normo-entendants (NE) les indices prosodiques dans un contexte discursif. Les adolescents IC ont en effet plus de difficultés que les adolescents NE à prévoir, compte-tenu des indices prosodiques contenus dans le contexte discursif, l'accent marquant le focus dans une phrase ; ces difficultés sont marquées par des temps de réaction plus longs chez les implantés cochléaires lorsqu'ils doivent détecter un phonème-cible contenu dans la phrase (Holt, Demuth, & Yuen, 2016). Par ailleurs, les adolescents implantés cochléaires n'utilisent pas les informations prosodiques de la même manière que les adolescents normo-entendants ; ils utilisent des contours mélodiques opposés à ceux utilisés par les adolescents normo-entendants lors de la production des mêmes actes de parole (e.g. contour montant plus important sur les requêtes d'informations que sur les directives chez les adolescents NE et contour montant plus important sur les directives que sur les requêtes d'informations chez les adolescents IC) (Holt, Yuen, & Demuth, 2017). Un dernier niveau de langage, relativement peu exploré en surdité, est le niveau pragmatique qui peut être défini comme "une habileté à utiliser la parole et les gestes de manière appropriée en tenant compte du contexte et des besoins de son interlocuteur" (Stephens & Matthews, 2014, p.14). D'une manière générale, les enfants sourds présentent, comme pour les autres niveaux de langage des niveaux de performances hétérogènes (Bebko, Calderon, & Treder, 2003). Tye-Murray (2003) a spécifiquement analysé, via des conversations filmées, les capacités des enfants implantés cochléaires, âgés entre 8 et 9 ans, à conserver la fluidité d'une conversation. Leurs résultats montrent, que comparé aux enfants normo-entendants, les enfants implantés présentent plus de silences et de bris de communication lorsqu'ils conversent avec un adulte inconnu, en l'occurrence ici une orthophoniste (Tye-Murray, 2003). En outre, lorsqu'ils conversent avec un adolescent normo-entendant du même âge qui leur est familier, les adolescents implantés cochléaires présentent des capacités conversationnelles identiques aux adolescents NE en quantité mais différentes en qualité : les adolescents IC produisent des

actes de parole différents de ceux produits par les adolescents NE lorsqu'ils conversent entre eux (i.e. plus de requête en clarification, moins de requête en confirmation et en élaboration) (Ibertsson, Hansson, Maki-Torkko, Willstedt-Svensson, & Sahlen, 2009). Les auteurs soulignent que ces scores sont à interpréter dans le contexte de la tâche qui est une tâche de communication référentielle, réalisée au calme, avec un adolescent normo-entendant connaissant l'adolescent sourd donc possiblement très adapté. Dans une autre tâche de communication référentielle (*map task*) réalisée avec un adulte étranger, les adolescents implantés cochléaires prennent moins la parole comparé aux adolescents normo-entendants, ils produisent moins d'acte de discours et marquent moins les frontières d'énoncés (Holt et al., 2017). Most et collaborateurs (2010) ont comparé, grâce à une grille d'analyse, le profil pragmatique d'enfants sourds implantés et appareillés avec des prothèses conventionnelles, à ceux d'enfants normo-entendants. Les résultats montrent un profil similaire entre enfants porteurs d'implants cochléaires et enfants porteurs de prothèses conventionnelles mais surtout une utilisation inappropriée des compétences communicatives, verbales et gestuelles, comparé à leurs pairs normo-entendants (Most, Shina-August, & Meilijson, 2010). Paatsch & Toe (2013) ont utilisé un contexte plus écologique que les études précédentes ; les auteurs ont mesuré les habiletés pragmatiques des enfants implantés cochléaires et appareillés avec des prothèses auditives conventionnelles (PC) à partir d'enregistrements de vidéos de conversations informelles entre les enfants sourds et leurs copains de classe normo-entendants, âgés entre 8 et 12 ans. Ils ont ensuite comparé ces mesures avec les mesures effectuées à partir des conversations des mêmes enfants normo-entendants mais entre eux. Les résultats montrent que lorsque les enfants sourds (avec PC et IC confondus) conversent avec un enfant normo-entendant, ils initient plus de sujet et présentent des tours de parole plus longs que leur partenaire contrairement aux paires d'enfants normo-entendants pour lesquelles les tours en termes d'initiation de sujet et de longueur, sont très bien équilibrés (Toe & Paatsch, 2013). La tendance des enfants implantés cochléaires à "dominer" la conversation, à moins se mettre à la portée de leur partenaire de communication, se retrouve dans les stratégies différentes qu'ils utilisent pour réparer les bris de communications lors d'une tâche de communication référentielle avec un inconnu. Dans cette tâche, les adolescents possèdent le tracer d'un itinéraire mais pas exactement la même carte que celui de leur partenaire. A partir de ce tracer, ils doivent guider leur

partenaire inconnu afin qu'il puisse retrouver sa route. Comparé à leurs pairs normo-entendants qui vont rechercher l'information manquante chez leur partenaire pour l'aider à avancer lorsque ce dernier est perdu, les adolescents implantés cochléaires ont plutôt tendance à répéter les informations de direction qu'ils viennent de donner sans utiliser des stratégies qui leur permettraient de combler les informations manquantes chez leur partenaire (Holt et al., 2017).

Pour résumer, les enfants implantés cochléaires présentent des niveaux de langage supérieurs à ceux de leurs pairs utilisant des prothèses conventionnelles. Implantés avant 18 mois (ou 24 selon les études), ces enfants peuvent obtenir des résultats comparables à ceux de leurs pairs normo-entendants à des tests de langage en expression et compréhension. Grâce à l'implantation précoce, ces enfants bénéficieraient en effet d'une plus grande aptitude à apprendre le langage de manière incidente. Les acquisitions les plus lentes portent sur les domaines de la grammaire, de la syntaxe et la perception et production des informations morphologiques. Cependant, ces enfants présentent aussi des différences (par rapport aux normo-entendants) dans le traitement de la prosodie du langage et les compétences discursives et conversationnelles. La communication auditivo-verbale versus la communication totale (i.e. comprenant l'utilisation de la langue orale associée à l'utilisation de la langue des signes) semble avoir une influence positive sur les capacités de production du langage des enfants IC, analysées à partir de corpus de langage spontané. Cependant, ces études ne permettent pas de déterminer si les enfants ayant besoin d'une communication totale ne possèdent pas, au préalable, un déficit ou des processus de traitement différents des enfants n'ayant besoin que de la communication auditivo-verbale. Malgré ces résultats favorables à l'implantation et en particulier l'implantation précoce, les auteurs mettent en garde contre l'interprétation des moyennes des scores obtenus aux tests de langage en soulignant l'importante erreur standard relevée dans les données (voir aussi van Wieringen & Wouters, 2015) ainsi que les retards subsistants, pour plus de la moitié de la population des enfants implantés cochléaires, vis-à-vis des capacités linguistiques des enfants normo-entendants (cf. l'article de revue de Ganek, McConkey Robbins, & Niparko, 2012). Bien qu'environ 50 % de la variabilité des résultats obtenus par les enfants implantés cochléaire aux tests de langage puissent être expliqués par l'âge d'implantation (Geers, 2009 ; van

Wieringen & Wouters, 2015), la différence d'organisation cérébrale et des processus de traitement retrouvés chez les enfants ayant souffert d'une déprivation auditive depuis leur naissance voire in utéro, pourraient expliquer, en partie, cette variabilité.

3. Conséquences de la surdité profonde congénitale sur le développement de l'organisation et du fonctionnement des structures cérébrales et des processus de traitement de l'information sensorielle

Plusieurs études sur l'animal et sur l'homme, réalisées en imagerie (en tomographie par émission de positron : PET scan) et en électrophysiologie (EEG), montrent que l'absence de stimulation auditive engendre une réorganisation cérébrale fonctionnelle (Lomber, Meredith, & Kral, 2010; Ponton & Eggermont, 2001). Le développement des aires cérébrales sensorielles, pré-existantes à la naissance, et les liens fonctionnels qui les relient, reposent en effet sur des périodes dites "sensibles" lors desquelles la création et le renforcement des synapses entre les neurones sont maximal. Si lors de ces périodes, la structure corticale ne reçoit un input sensoriel suffisant, en l'occurrence ici, si le cortex auditif ne reçoit pas de stimuli acoustiques, le développement de l'aire auditive primaire - aire qui traite les caractéristiques temporelles et fréquentielles des sons - et ses connexions avec les aires secondaires, qui attribuent des représentations aux sons, se désorganisent. Cette altération des connexions entre les aires auditives primaires et les aires de plus haut niveau de traitement ne permet pas la mise en place de voies cortico-fugales qui partent des cortex secondaires associatifs vers les cortex primaires et les aires sous-corticales (voir figure 4.2). Le développement de ces voies est important car elles sont la source de processus dits "*top-down*" qui modulent la perception de stimuli auditifs que nous recevons. On sait par ailleurs que plus l'expérience auditive est importante, plus la modulation des stimuli auditifs sera rapide et précise (Lesicko & Llano, 2017). Le manque de modulation des représentations de haut niveau lors de la perception (processus *top-down*) altère les capacités de catégorisation et d'organisation des informations auditives nouvelles (Kral & Eggermont, 2007). Ainsi, même si les facteurs génétiques déterminent les connexions potentielles entre les structures cortico-sous-corticales et cortico-corticales, c'est bien la qualité des stimulations et de l'expérience auditive qui va par la suite façonner et établir une stabilité de ces connexions,

source de processus de traitement efficaces du langage (voir article de revue Kral, Yusuf, & Land, 2017).

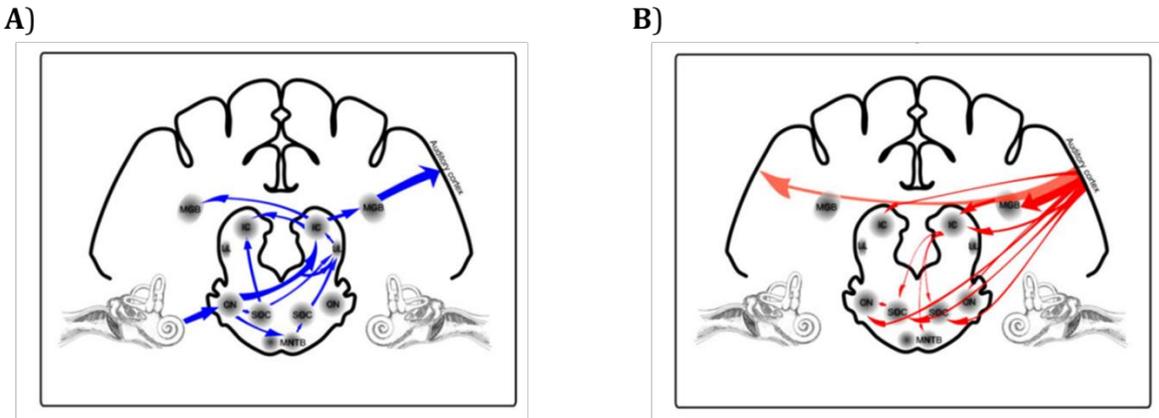


Figure 4.2. Représentation des voies auditives afférentes - bottom-up - et cortico-fugales - top-down extraite de Kral et Eggermont (2007). Panel A) Depuis le nerf auditif, l'influx nerveux est transmis aux différents relais du tronc cérébral et arrive au cortex auditif primaire puis secondaire. Panel B) Depuis le cortex auditif, l'influx nerveux redescend vers le tronc cérébral pour moduler la réponse nerveuse en périphérie.

Plus précisément, les études révèlent différents marqueurs qui signent l'impact d'une déprivation auditive sur l'organisation fonctionnelle du système auditif.

Ainsi, la latence et la forme de l'onde P1, qui reflètent la durée et la qualité de la propagation de l'information électrique du tronc cérébral au cortex auditif primaire - autrement dit la maturation des voies auditives - est considérablement retardée et déformée chez les enfants implantés cochléaires après 7 ans (Sharma, Dorman, & Spahr, 2002a, 2002b; Sharma, Spahr, Dorman, & Todd, 2002). Contrairement aux enfants implantés avant 3,5 ans, chez les enfants implantés tardivement, la latence de l'onde P1, qui apparaît, à la naissance, 300 ms post-stimulation puis passe à 100 ms à 2 ans et enfin entre 50 et 70 ms à l'âge adulte, ne parvient jamais à atteindre celle des enfants normo-entendants (Sharma, Gilley, Dorman, & Baldwin, 2007) (voir figure 4.3).

Alors que l'onde P1 est générée dans le cortex auditif primaire, la source du complexe N1/P2 se trouverait dans le cortex auditif secondaire, de plus haut niveau. Le développement

de ce complexe N1/P2 est associé à une augmentation du couplage cortico-cortical (entre les aires auditives primaires et les aires auditives secondaires) et une augmentation du traitement auditif et du langage (voir Chapitre 3 Eggermont & Moore, Human Auditory Development, 2012). Alors que les enfants implantés avant 3,5 ans développent ce complexe N1/P2 à l'adolescence comme les adolescents normo-entendants (voir figure 4.3 panel K) et panel B) respectivement), les enfants implantés tardivement ne développent jamais ce complexe (voir figure 4.3 panel L). Le découplage entre les aires auditives primaires et secondaires chez ces enfants pourrait expliquer leurs difficultés à attribuer du sens aux perceptions reçues via l'implant cochléaire. En outre, comme l'expliquent Sharma et al. (2015), le découplage entre les cortex auditifs primaires et secondaires va engendrer une réorganisation corticale : les inputs sensoriels provenant de la modalité visuelle et somatosensorielle vont "coloniser" les voies reliant les aires auditives primaires et secondaires et c'est ainsi que les cortex auditifs secondaires vont être activés chez des adultes sourds, lors de la présentation de stimuli visuels ou somatosensoriels. La réorganisation des voies fonctionnelles existant à l'intérieur et entre les cortex sensoriels semble corrélée aux résultats retrouvés en langage même chez des enfants implantés précocement (Sharma, Campbell, & Cardon, 2015). Lors d'une récente étude, Feng et collaborateurs (2018) ont développé un modèle permettant de prédire les scores de langage des enfants futurs implantés cochléaires. A partir des données IRM pré-implantation, les auteurs ont comparé les données neuroanatomiques des enfants candidats à l'implantation à celles d'enfants normo-entendants du même âge. Ils ont ainsi pu observer les différences de densité de matière grise et blanche et de réseaux fonctionnels entre les deux populations et établir les régions qui avaient été affectées par la déprivation sensorielle. Les auteurs ont également mesuré l'amélioration de la perception auditive et les compétences linguistiques des enfants avant l'implantation et à 6 mois post-implantation. Dans ce modèle, l'âge, le sexe, la classe socio-économique des parents, la latéralisation de l'IC, l'audition résiduelle et les différences et similitudes neuroanatomiques entre enfants sourds et enfants normo-entendants ont été entrées comme prédicteurs des capacités perceptives et linguistiques des enfants futurs candidats à l'implant. Les résultats montrent que ce ne sont pas les régions affectées par la déprivation sensorielle qui prédisent le mieux l'amélioration de la perception et des

compétences linguistiques après implantation mais les régions cérébrales qui sont restées identiques à celles des enfants normo-entendants malgré la déprivation auditive.

Autrement dit, ce n'est pas l'altération des régions auditives primaires mais la préservation des régions auditives et cognitives de haut niveau telles que le gyrus temporal supérieur, les cortex fronto-pariétal et la voie dorsale, reliant les structures auditives aux structures motrices, qui prédit le mieux les scores de langage des enfants implantés cochléaires. Ce réseau cognitif est en effet connu pour être hautement impliqué dans la perception de la parole (cf. partie 1 et 2) mais aussi dans le traitement séquentiel de l'information sensorielle.

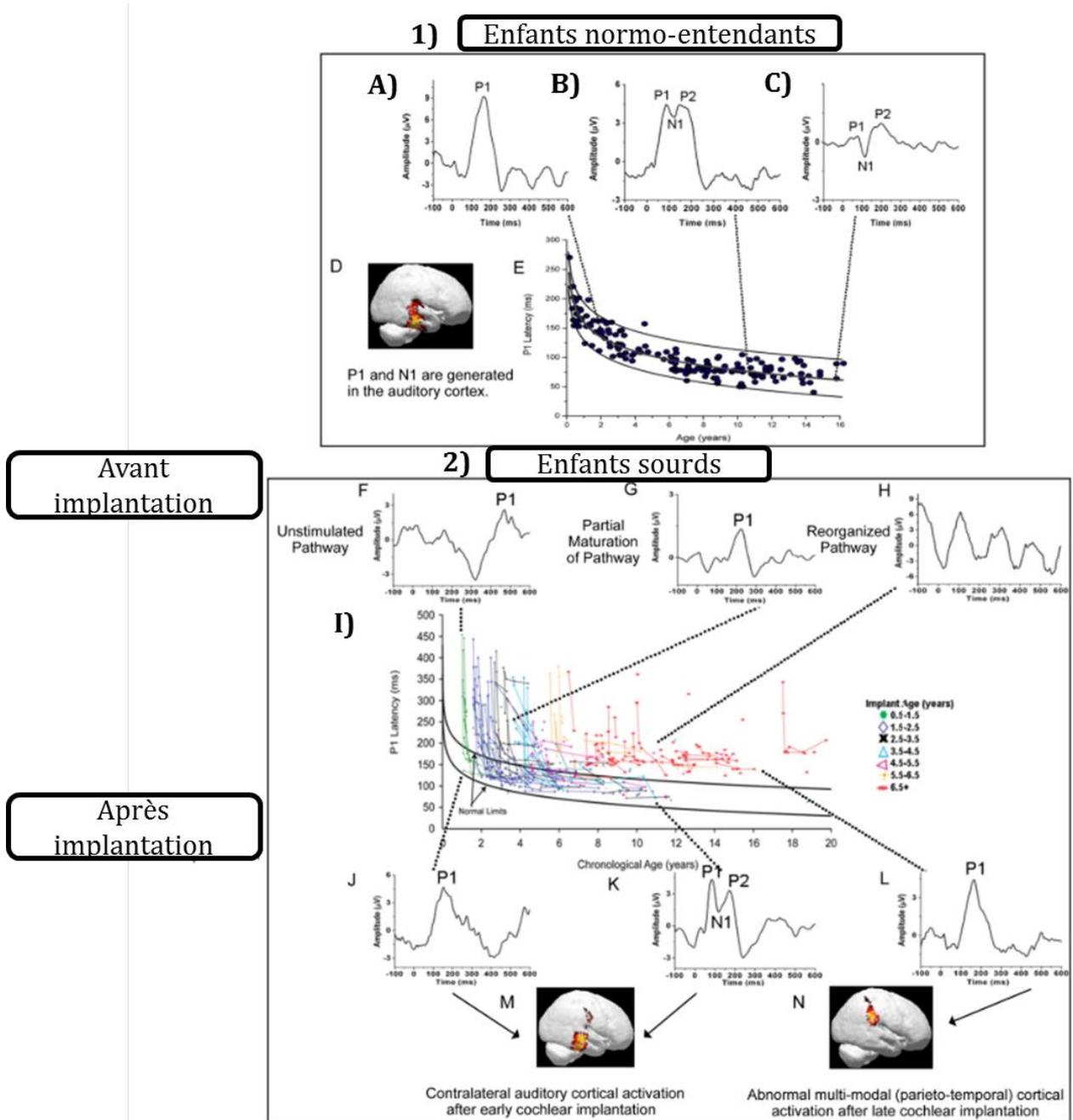


Figure 4.3. Représentation du développement de l'onde P1 et du complexe P1/N1/P2 chez les enfants normo-entendants et chez les enfants sourds en fonction de l'âge d'implantation extraite de Sharma et Kral (2012). Graphique n°1 : développement de l'onde P1 et du complexe P1/N1/P2 en fonction de l'âge chez les enfants normo-entendants. Les panels A), B) et C) représentent l'évolution de la maturation de ces différentes ondes. Graphique n°2 : Panel I) Les lignes noires représentent les limites des latences des enfants

normo-entendants en fonction de l'âge. Les points de couleurs reliés par des lignes représentent le développement de la latence de l'onde P1 d'enfants implantés cochléaires à différents âges (voir la légende des âges sur la droite). On peut voir qu'avant 1, 5 ans ou 2 ans, tous les enfants implantés cochléaires, quel que soit l'âge d'implantation, ont des latences retardées, i.e. latences qui vont au-delà des limites de latences retrouvées chez les enfants normo-entendants. Mais les points verts, violet et noir, qui correspondent aux enfants implantés avant 3,5 ans, sont situés dès 6 à 8 mois post-implantation à l'intérieur de ces lignes ce qui signifie que ces enfants implantés précocement sont parvenus à rattraper les latences des enfants normo-entendants contrairement aux enfants implantés plus tardivement représentés par les points d'autres couleurs.

Certains auteurs pensent en effet que la déprivation auditive n'affecte pas seulement la perception et la production du langage mais également des processus neurocognitifs plus généraux (Houston et al., 2012). Des études mettent en évidence des déficits dans la formation de concepts (Castellanos, Pisoni, Kronenberger, & Beer, 2015) ainsi que dans des tâches impliquant les fonctions exécutives (Beer et al., 2014; Figueras, Edwards, & Langdon, 2008; Kronenberger, Pisoni, Henning, & Colson, 2013). Une autre compétence, semble affectée chez les enfants sourds implantés cochléaires. Conway et collaborateurs (2009) ont en effet émis l'hypothèse que la déprivation auditive n'affecterait pas seulement la perception auditive mais que celle-ci aurait des conséquences sur des habiletés cognitives liées à l'apprentissage implicite de séquences (Christopher M. Conway, Pisoni, & Kronenberger, 2009). L'apprentissage implicite est par exemple cette capacité qui nous permet d'apprendre de manière incidente une succession de mouvements ou d'évènement dans un ordre précis. Cet apprentissage de patterns d'actions nous permet de développer une compétence comme apprendre à jouer d'un instrument ou faire des prédictions sur le déroulement des évènements qui ont lieu habituellement dans notre environnement. Plusieurs études ont montré que cette compétence cognitive, générale à plusieurs domaines d'apprentissage, serait nécessaire au développement des compétences linguistiques (voir Deocampo, Smith, Kronenberger, Pisoni, & Conway, 2018 pour une revue de la littérature). Pour tester la présence ou l'absence de cette capacité chez les enfants sourds, Conway et collaborateurs (2011) ont présenté à des enfants implantés cochléaires ainsi qu'à des enfants

normo-entendants une succession de stimuli visuels dont l'ordre d'apparition dépendait d'une grammaire artificielle. Dans une phase pré-test, les enfants ont appris de manière implicite l'ordre de présentation des stimuli. Dans la phase de test les enfants devaient rejouer les séquences apprises. Les auteurs ont analysé la précision du rappel des séquences à la fois lors de la présentation de séquences qui suivaient la grammaire artificielle mais aussi lors de la présentation de séquences qui suivaient un ordre aléatoire. Les résultats ont montré que les enfants sourds n'obtenaient pas de meilleures performances au rappel de séquences lorsque celles-ci étaient organisées selon une structure non aléatoire (i.e. qui suivaient les règles d'une grammaire artificielle) alors que les enfants normo-entendants ont bénéficié de cette organisation des séquences comparé à la condition où les séquences suivaient un ordre aléatoire. Les auteurs en ont conclu que les enfants ayant souffert d'une déprivation auditive présentaient des déficits d'apprentissage implicite de séquences dans toutes les modalités sensorielles (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Mais d'autres auteurs (Hall, Eigsti, Bortfeld, & Lillo-Martin, 2017; Torkildsen, Arciuli, Haukedal, & Wie, 2018) ont émis des réserves quant à la différence de l'effet retrouvé entre les deux populations et au type de paradigme utilisé. Hall et collaborateurs (2017) ont ainsi tenté de répliquer l'étude de Conway et collaborateurs (2011) mais sur trois populations différentes : des enfants sourds signants, des enfants sourds avec implants cochléaires et des enfants normo-entendants. Ils ont ensuite testé ces trois populations sur les mêmes compétences d'apprentissage implicite mais à partir d'une tâche de temps de réaction en série (Serial Reaction Time : SRT). La réplification de l'étude Conway (2011) ne montre pas les mêmes résultats : aucune des trois populations ne présente des capacités d'apprentissage implicite suite à la tâche basée sur l'apprentissage d'une grammaire artificielle. Les résultats obtenus suite à la tâche de SRT montrent par contre que les trois populations possèdent de robustes capacités d'apprentissage implicite et qu'il n'y a pas de différence entre les trois populations (Hall et al., 2017). von Koss Torkildsen et collaborateurs (2018) ont voulu répliquer ces études en utilisant une tâche d'apprentissage implicite impliquant, entre autre, moins de stratégies de rappel verbal que la tâche de Conway (2011) et les résultats montrent des capacités d'apprentissage implicite identiques chez les enfants implantés cochléaires et les enfants normo-entendants. Cependant, même si ces enfants développent des capacités d'apprentissage implicite identiques aux enfants normo-entendants, il semble néanmoins

que la déprivation auditive engendre des déficits d'attention précoces à la parole (à 1 et de 3 à 6 mois post-implantation) qui ne sont pas expliqués par les données démographiques habituellement corrélées telles que l'âge d'implantation, la quantité d'audition résiduelle ou le mode de communication (Wang, Shafto, & Houston, 2018).

Il semble donc difficile de déterminer avec précision la cause de la variabilité des résultats retrouvés dans les profils linguistiques des enfants implantés cochléaires, néanmoins, à la lumière des données d'imagerie cérébrale et électrophysiologiques il semble que les structures corticales de haut niveau, impliquées dans des opérations cognitives générales, soit en partie responsables des déficits perceptifs et linguistiques retrouvés chez ces enfants même implantés précocement. Cependant, comme expliqué aux chapitres 1 et 2, la pratique de la musique favorise la stimulation de plusieurs opérations cognitives en partie utilisées lors du traitement du langage mais également l'activation des aires cérébrales de haut niveau et notamment les connexions fonctionnelles entre les aires auditives et les aires motrices, impliquées dans l'attention et perception de la parole.

Nous allons montrer dans cette dernière partie que la musique peut avoir un effet sur la perception et la production du langage des enfants sourds.

4. Effet de la stimulation musicale sur la perception auditive et la production de la parole chez l'enfant sourd

Bien que la musique soit utilisée en rééducation par des orthophonistes musiciennes et non musiciennes à des fins de découverte des styles musicaux, des timbres des instruments, des différents paramètres sonores, de la mise en musique d'émotions jouées par des personnages de fictions, elle est encore très peu utilisée à des fins de rééducation du langage. Pourtant, de plus en plus d'études et d'articles de revue recommandent l'utilisation de la musique dans les prises en charge et en particulier dans la rééducation du langage (François, Grau-Sánchez, Duarte, & Rodriguez-Fornells, 2015; Fujii & Wan, 2014; Thaut, McIntosh, & Hoemberg, 2015). Quelques études analysant les effets de transfert d'un entraînement musical sur les compétences langagières des personnes porteuses d'un implant cochléaire commencent aussi à voir le jour. La plupart de ces études sont réalisées chez des adultes post

linguaux (voir par exemple Fuller, Galvin, Maat, Başkent, & Free, 2018) et quelques-unes commencent à être réalisées chez les enfants sourds congénitaux.

Deux études ont testé les effets d'un amorçage musical de type rythmique sur les compétences langagières des enfants sourds congénitaux implantés cochléaires. Une première étude de Cason et collaborateurs (2015) a analysé la qualité de la répétition de phrases chez un groupe d'enfants sourds implantés cochléaires et appareillés de manière conventionnelle directement après l'écoute et la reproduction d'une structure rythmique musicale. Ils ont comparé le pourcentage de voyelles, consonnes, syllabes et mots répétés sans l'amorçage et après l'amorçage ; la structure métrique de cet amorçage pouvant correspondre ou pas à la structure métrique des phrases à répéter. Les résultats montrent que d'une manière générale, les enfants sourds obtiennent de meilleures performances de répétition après l'amorçage versus aucun amorçage et ce, que l'amorce soit identique ou pas à la phrase en termes de structure métrique. Cependant, cet effet est retrouvé plus important chez les enfants implantés cochléaires comparé aux enfants appareillés de manière conventionnelle (Cason, Hidalgo, Isoard, Roman, & Schön, 2015). Selon le même type de protocole expérimental, Bedoin et collaborateurs (2017) ont testé, chez des enfants sourds congénitaux implantés cochléaires, les effets d'un entraînement morphosyntaxique après un amorçage musical de type rythmique (8 sessions de 20 minutes) et après un amorçage auditif à partir de sons environnementaux sans structure rythmique (8 sessions de 20 minutes). Les auteurs ont réalisé des mesures de compréhension syntaxique et de jugement grammatical ainsi que des mesures de répétition de non-mots, d'attention visuo-spatiale et de mémoire avant les entraînements, après le premier entraînement et après le deuxième entraînement. Les résultats montrent que quel que soit le type d'amorçage qui précède l'entraînement, les enfants implantés cochléaires, âgés de 7 ans en moyenne, améliorent leurs performances dans les jugements grammaticaux, le traitement de la syntaxe, la répétition de non-mots et l'attention mais que la différence de performances dans les jugements grammaticaux et la répétition de non-mots est plus importante après l'entraînement précédé d'une amorce musicale qu'après celui précédé d'une amorce auditive (Bedoin et al., 2017). Les deux études décrites ci-dessus ont mesuré les effets d'un entraînement musical à court terme (amorçage)

sur les compétences langagières, mais d'autres études ont étudié les effets d'un entraînement musical de plus longues durées.

Torppa et collaborateurs (2014) ont par exemple évalué de manière indirecte, les effets d'une pratique musicale régulière sur les compétences linguistiques des enfants implantés cochléaires. Dans cette étude, les auteurs analysent trois groupes d'enfants : un groupe d'enfants normo-entendants, un groupe d'enfants implantés cochléaires et un groupe d'enfants implantés pratiquant une activité musicale à l'école, dans une association ou ayant pratiqué une activité musicale avec ses parents à un plus jeune âge. Ils comparent ces trois groupes d'enfants sur une tâche de discrimination d'accents dans les mots, de discrimination de paramètres acoustiques (durée, intensité et fréquence fondamentale) et sur une tâche de mémoire de travail. Les résultats montrent que les enfants implantés ayant suivi une activité musicale ont de meilleurs résultats que ceux qui pratiquent une autre activité mais surtout qu'ils obtiennent des performances équivalentes à celles des enfants normo-entendants dans les tâches de discrimination de la F0, de la durée, de répétition de chiffres (i.e. mémoire de travail) et dans la tâche de perception de la prosodie (Torppa, Faulkner, et al., 2014).

Dans une méta-analyse récente Gfeller (2016), a recensé les quelques études qui ont évalué plus directement les effets de la pratique de la musique sur les capacités musicales et linguistiques des enfants implantés cochléaires. Selon l'auteur, il est difficile d'appréhender la taille des effets générés par les interventions car les méthodologies employées (contenu des stimulations : activités multimodales versus spécifiques, format des stimulations : approches pédagogiques holistiques en groupe versus individualisées et/ou à partir de programmes informatisés, fréquence et durée des stimulations : de 2 semaines à 2 ans), l'âge des participants, l'âge et la durée d'implantation, le type d'implant utilisé et la manière dont ces effets sont reportés sont très différents d'une étude à l'autre. Les critères de sélection de ces études à savoir : publication dans des revues à comité de lecture, écrites en anglais, réalisée chez des enfants de moins de 18 ans, ont permis d'en recenser seulement neuf ! En outre, la plupart des résultats des études sont basés sur des analyses de corrélations entre l'amélioration des capacités perceptives et/ou linguistiques et la durée de l'entraînement musical (voir par exemple Chen et al., 2010) (Gfeller, 2016). Cependant, nous présenterons ici seulement les résultats de quelques rares études qui incluent un groupe contrôle dans

leurs analyses ; en effet, seules ces études nous permettent de conclure précisément sur les effets de transfert d'un entraînement musical sur les capacités linguistiques.

Rochette et collaborateurs (2014) ont testé chez des enfants sourds profonds implantés et appareillés avec des prothèses conventionnelles, les effets d'un entraînement musical hebdomadaire (1 h00 pendant 2,6 ans) sur leurs capacités de perception auditive et compétences cognitives nécessaires à la perception du langage. Les résultats, qui comparaient les performances d'un groupe ayant suivi l'entraînement musical et un groupe sans entraînement musical, montrent que les enfants qui ont suivi l'entraînement musical ont de meilleures performances dans la tâche d'analyse de scènes auditives - qui est une tâche qui mobilise des capacités nécessaires à la discrimination dans le bruit - ainsi que de meilleurs résultats dans la tâche de mémoire auditive et de discrimination phonétique (Rochette, Moussard, & Bigand, 2014).

Fu et collaborateurs (2015) ont également montré que comparé à un groupe d'enfants implantés cochléaires non entraînés, ceux qui suivent un entraînement auditif quotidien (30 minutes par jour pendant 10 semaines) basé sur la reconnaissance de contours mélodiques, ont de meilleures capacités à identifier des contours mélodiques nécessaires à la perception de la prosodie du langage (Fu, Galvin, Wang, & Wu, 2015). Good et collaborateurs (2017) ont quant à eux testé chez 18 enfants implantés cochléaires âgés de 6 à 15 ans, les effets d'un entraînement de piano versus des cours de peinture d'une durée de 6 mois chacun (à raison de 30 minutes par semaine), sur les compétences musicales et linguistiques. Les mesures ont été récoltées à trois temps différents (en pré, milieu et post entraînement/cours). L'évaluation des aptitudes musicales mesurées à l'aide de la Montréal Battery for Evaluation of Amusia portaient sur la capacité à discriminer des contours mélodiques, des intervalles, des échelles, des rythmes ainsi que sur la capacité à mémoriser une mélodie de manière incidente (i.e. analyser dans quelle mesure le sujet, alors qu'il réalise une autre tâche, est capable de mémoriser une mélodie sans que cela lui ait été explicitement demandé). Les aptitudes langagières portaient sur la capacité à reconnaître, suite au visionnage (modalité audio-visuelle) ou à l'écoute seule (modalité auditive pure), des émotions véhiculées par la prosodie avec différents niveaux d'expressivité. Les résultats ont montré que l'entraînement musical, comparé à l'activité peinture, permettait l'amélioration de la discrimination des

contours mélodiques, des rythmes, de l'apprentissage incident des mélodies et des scores au test de prosodie émotionnelle des enfants implantés cochléaires (Good et al., 2017).

Pour résumer, il existe peu d'études à notre connaissance qui ont analysé les effets d'une stimulation musicale sur les capacités de production de la parole chez l'enfant sourd. Néanmoins, il semble qu'un entraînement musical non seulement améliore chez les enfants implantés cochléaires, la perception des paramètres acoustiques musicaux tels que la hauteur et les contours mélodiques, éléments présents dans la prosodie du langage, mais que la pratique de la musique engendre également des effets de transfert direct sur les compétences cognitives et linguistiques nécessaires à une bonne perception du langage.

Partie Expérimentale

Le rythme de parole est un paramètre essentiel à prendre en compte pour la compréhension du message oral et tout particulièrement en situation de conversation. Pour percevoir le rythme et pouvoir l'utiliser de manière optimale dans la conversation, il faut cependant pouvoir développer des processus de traitement temporels de haut niveau similaires à ceux utilisés dans le traitement de la musique.

Chez l'enfant normo-entendant, le développement des capacités de traitement temporel semble passer par les stimulations auditives reçus in utéro et en période prélinguistique. Or, les enfants sourds n'ont pu bénéficier de ces stimulations au moment où la plasticité cérébrale est optimale pour le développement des voies auditives et du langage et souffrent, malgré le port d'un implant cochléaire ou d'une prothèse conventionnelle, de difficultés de perception et de compréhension du langage plus spécifiquement en situation de conversation à plusieurs.

Durant ces trois années de thèse, nous avons cherché à mesurer les capacités temporelles musicales et linguistiques des enfants sourds en particulier des enfants implantés cochléaires. Nous avons également cherché à analyser les effets d'une stimulation musicale rythmique sur leurs capacités de perception et de production de la parole en situation d'interaction en proposant la tâche expérimentale de parole après une séance de stimulation linguistique/auditive et après une séance de stimulation rythmique active, chacune d'une durée de 30 minutes.

Ainsi, dans une première expérience, nous avons comparé les capacités rythmiques des enfants sourds et des enfants normo-entendants dans des tâches de synchronisation sensori-motrice sur des stimuli rythmiques différents d'un point de vue de leur structure acoustique et temporelle. Dans deux autres expériences, nous avons utilisé un nouveau paradigme dans lequel les enfants normo-entendants et les enfants sourds devaient dénommer des images en alternance avec un partenaire. Durant cette alternance nous avons manipulé dans une expérience la vitesse de l'alternance et la rythmicité des mots échangés. Nous avons pu ainsi mesurer les capacités des enfants sourds à demeurer consistant et précis dans l'alternance malgré les variations de la parole du partenaire virtuel autrement dit leurs capacités à s'accommoder temporellement à son rythme de parole. A partir du même

paradigme, dans une deuxième expérience, nous avons également manipulé la régularité de l'alternance en introduisant des déviations temporelles dans la régularité de l'alternance imposée par le partenaire virtuel. Nous avons pu alors mesurer, grâce à des mesures électrophysiologiques, la sensibilité de leur système perceptif à des déviations de la régularité dans l'alternance ainsi que leurs capacités comportementales d'adaptation à ces déviations. Pour ces deux expériences, nous avons également, mesuré les effets de la stimulation rythmique versus linguistique/auditive.

Expérience n°1

Musical Rhythmic Abilities of Children with Cochlear Implant

Thanks to neonatal screening and early implantation, deaf children with a cochlear-implant (CI), educated in an oralising environment, can reach receptive vocabulary levels close to those of their normal-hearing peers (Hayes et al., 2009) and superior to those of their peers wearing conventional hearing aids (Baldassari et al., 2009; Truy et al., 1998). However, some linguistic difficulties persist compared to normal-hearing children. For instance, in terms of verbal fluency at both phonological and semantic levels (Wechsler-Kashi, Schwartz, & Cleary, 2014) but also in verbal reasoning and syntactic processing (Geers, Moog, Biedenstein, Brenner, & Hayes, 2009). Quite common are difficulties with lexical comprehension and morphosyntactic production-comprehension, especially when requiring the discrimination of words that are close from a phonological point of view (Caselli, Rinaldi, Varuzza, Giuliani, & Burdo, 2012).

Syntax deficits appear to have a particularly negative impact on the metacognitive skills of children with hearing impairment (HI) such as the development of the theory of the mind or reading comprehension (Lederberg, Schick, & Spencer, 2013). An analysis of language levels in adolescent implanted between the ages of 2 and 5 reveals persistent difficulties with higher level language skills than those usually tested: even though these adolescents have a good understanding of isolated words and of concrete statements, they have lower scores in discursive and abstract reasoning tasks (Geers & Sedey, 2011). In addition, the treatment of prosodic features is also impaired in cochlear implanted children. They exhibit deficits in discriminating vocal emotions possibly due to a manifest limitation in pitch perception (Torppa et al., 2014) and intonation contours (Jiam, Caldwell, Deroche, Chatterjee, & Limb, 2017). Compared to their normal-hearing peers, they are also less attentive to the prosodic cues that can be highly relevant to discourse comprehension (Holt, Yuen, & Demuth, 2017). Considering the strong link between attention to speech and speech perception in children with CI (Wang, Shafto, & Houston, 2018), one important challenge of future research is to define a re-educative approach allowing to better focus attention to the most relevant clues of speech.

A potential candidate for this challenging task is the use of rhythmic-based interventions.

According to the Dynamic Attending Theory (DAT), the level of attention paid to an auditory stimulus is guided by a synchronization phenomenon between the temporal envelope of the stimulus and the oscillatory activity of the auditory system. This phenomenon, called entrainment, would occur at the different temporal scales present in both music and speech. Thus, the auditory system could be coupled to the frequency of occurrence of sounds or syllables - i.e. short time scale - and to longer

time scales such as metric structures or prosodic units. The regularity of the appearance of temporal events in music and speech can be exploited by the listener to predict the occurrence of future events and thus better perceive them.

Speech does not present a clear-cut periodicity. In other words, speech units are not isochronously distributed. Nonetheless, several studies have shown that listeners are able to tap a rhythm when they hear speech; more precisely, listeners regularly tap at times corresponding to perceptually anchor points (p-center) within each accented syllable. In other words, perception of p-center distribution is regularized (Lehiste 1977, Fowler, 1979). This regularization of time intervals would enable listeners to better anticipate the moment of occurrence of the next accent.

On the other hand, speech, like music, has different levels of hierarchically organized accents or prominence (Di Cristo, 2002; (Jun & Fougeron, 2002). The synchronization of the perceptual system on the different levels of the hierarchy would increase the perception of regularity as well as the quality of sensory processing (Jones, 2016; London, 2004). Because the rhythm of speech and the musical rhythm seem rather similar in terms of organization and processing (Fitch & Martins, 2014; Haegens & Zion Golumbic, 2018), some authors have used musical rhythm as a tool of rehabilitation in populations suffering from language disorders.

Thus, rhythmic training over short, medium or long periods would improve speech processing at different processing levels in clinical populations (Schön & Tillman, 2015). Sentence perception and production improves following rhythmic priming in children with hearing impairment wearing CI or hearing aids (Cason, Hidalgo, Isoard, Roman, & Schön, 2015). The syntactic understanding of spoken sentences improves following a short stimulation using rhythmic sequences in children with specific language disorders (Bedoin, Brisseau, Molinier, Roch, & Tillmann, 2016; Przybylski et al., 2013). The temporal adaptation capacities of children with CI in a situation of verbal interaction improve following 30-minute rhythmic intervention compared to a standard speech therapy session (Hidalgo, Falk, & Schön, 2017). The phonological awareness improves in children with dyslexia following a one-year rhythmic musical intervention (Flaugnacco et al., 2015).

However, rhythmic skills are not a uniform but rather multidimensional component, linked to different cognitive abilities (Tierney & Kraus, 2015). Thus, depending on whether we process short or long speech time intervals (see (Rosen, 1992) for the different temporal scales of speech), different neural networks may be at work (Poeppel, 2003, Giraud & Poeppel, 2012; Luo & Poeppel, 2012). Thus, it seems relevant to gather a better understanding of what rhythmic skills are preserved or altered in

children with CI as well as to bridge these skills to speech and language abilities, as this will allow a better targeting of the therapeutic intervention.

To our knowledge, while a few studies have investigated music perception of cochlear implanted children (Hopyan et al., 2012; Innes-Brown et al., 2013; Roy et al., 2014), none has specifically addressed the rhythmic production abilities. Our current aim was to fill this gap using a battery of different rhythmic tests to evaluate several types of rhythmic skills from sensory-motor skills to more integrated skills such as reproducing a complex rhythm. The performance of cochlear implanted deaf children in these different skills is compared to an age-matched control population and interpreted relative to working memory and language abilities.

Population

Thirty-two children aged from 5 to 10 years ($mean = 88.2$ months, $SD = 19.8$ months) with severe to profound hearing loss, without any known additional disorders were recruited from the pediatric implantation center of Lyon (Edouard Herriot hospital, France) and from four independent speech therapy centers in Marseille (France). These children wore one or two cochlear implants and/or a contralateral hearing aid (see table 1 for more details).

Twenty-four children with normal hearing (NH) were recruited from two *elementary schools*: one in Lyon and one in Marseille (France). These children were all French native speakers without any known visual, speech, cognitive or hearing disorder and were paired in age with children with hearing loss (HL) ($mean = 90$ months, $SD = 21.2$ months, $t(54) = -.33$, $p = .74$). This experiment was approved by the Ethics Committee Sud Méditerranée I (n°ID RCB: 2015-A01490-49). All parents signed a consent form.

Procedure

All children were assessed with three rhythmic tests and a working memory test. Children with HL were also assessed with a language test. Children were seated in front of the experimenter (a speech therapist) in a quiet room. In the rhythmic production tasks, they were asked to tap with their dominant hand (children > 8 years old) or with a plastic frog (children < 8 years old) either on a digital tablet (I.T. Works 10.1, 25.65 cm) or onto a cardboard box containing a ZOOM H4n digital recorder. All sounds were delivered by loudspeakers Creative Inspire T10 and the volume was adapted according to each child's hearing level. A training phase with the speech therapist preceded each task. The experimental session lasted ~30 minutes.

Rhythmic tests

The paced tapping test was taken from the BAASTA battery (Dalla Bella et al., 2016). Children were asked to tap in synchrony with 60 isochronous piano tones (frequency = 1319 Hz) (Inter Trials Intervals 600 ms). Tap onsets were calculated using the BAASTA algorithm. Analyses were run on the consistency in the children tapping behavior.

The tapping to music test was taken from the Beat Assessment Test (BAT, Iversen & Patel, 2008). Children were asked to tap in synchrony with the beat of one jazz (*Tuxedo Junction* of Glenn Miller) and one pop (*A Chorus Line* of Boston Pops) musical excerpt. Each excerpt was repeated twice each. Tap onsets were obtained using Adobe Audition 2.0 according to a semi-automatic procedure and analyses were run with a custom Matlab script on the same measure as the paced tapping task.

The tapping on complex rhythmic sequences was taken from Tierney & Kraus (2015). In this task children listen to a complex rhythm that is repeated eight times in a row and they are asked to tap along. Only the performance on the last repetition was taken into account, since this is typically the best and most stable performance and it is less dependent on working memory skills. Rhythm performances were scored by three independent judges from 1 to 9 depending on the performance similarity to the stimulus (1 no reproduction, 9 perfect performance); the final mark for each child was the average of the 4 scores.

The working memory test was selected from the WISC IV (indirect span). Children were asked to repeat sequences of digits of increasing length in the reverse order. The test stops when the child gives a wrong answer on more than 2 items of a given length in a row.

The language test was only assessed in children with HL because children with NH would be at ceiling. The task consisted of repeating short sentences. Twenty sentences were selected from Cason and collaborators (2015), composed of six syllables. Children listened to each sentence twice and were then asked to repeat it. Repetition performances were scored by a speech therapist. One point was attributed for each correctly repeated syllable; the final mark was the average for the 20 sentences which was then converted in percentage.

Analysis

All statistical analysis were computed in R version (R Core Team, 2013). Because data were distributed asymmetrically we conducted Mann-Whitney U tests between groups. Consistency in the tapping behavior was assessed by the vector length in the circular space. This variable was logit transformed for statistical analyses (Falk, Müller & Dalla Bella, 2015). When appropriate, model comparison was realized using the Akaike Information Criterion (AIC).

Results

Children with CI tapped less consistently with the metronome (paced tapping) and with the different musical excerpts (music tapping) compared to NH children (respectively $U = 219, p < .01, d = .70$ and $U = 171, p < .001, \text{Cohen's } d = 1.04$) (see figure X, panel A). They also performed more poorly than children with NH in reproducing complex rhythmic sequences ($U = 131, p < .001, d = 1.39$) (see figure X, panel B) as well as in the working memory task ($U = 228, p < .05, d = .65$, see figure X, panel C). Even when considering their working memory capacities, children with CI's complex rhythmic reproduction capacities are worse than NH children performances ($\beta = 0.48, SE = 0.14, t = 3.49, p < .01$).

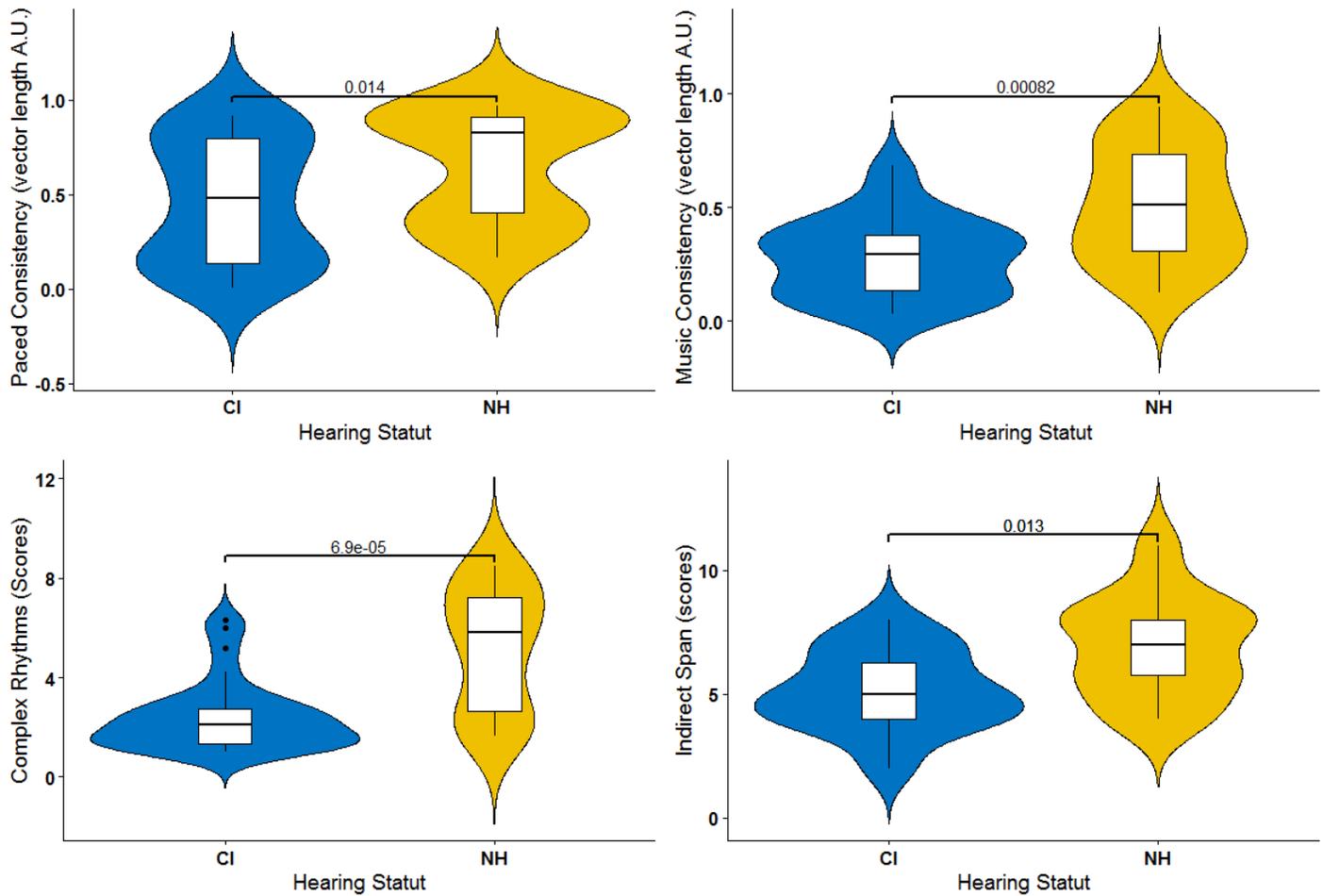


Figure X Panel A), B), C), D)

We then computed two correlational analyses with Spearman's ranks. The first was computed on data of all children (with CI and with NH) using paced tapping, music tapping, complex rhythm reproduction, indirect span and age as dependent variables. Results are presented in figure X. As expected performance in the tapping and rhythmic tasks are positively correlated. Also working memory correlates with the temporal tasks, while age does not seem to be strongly related to task performance overall.

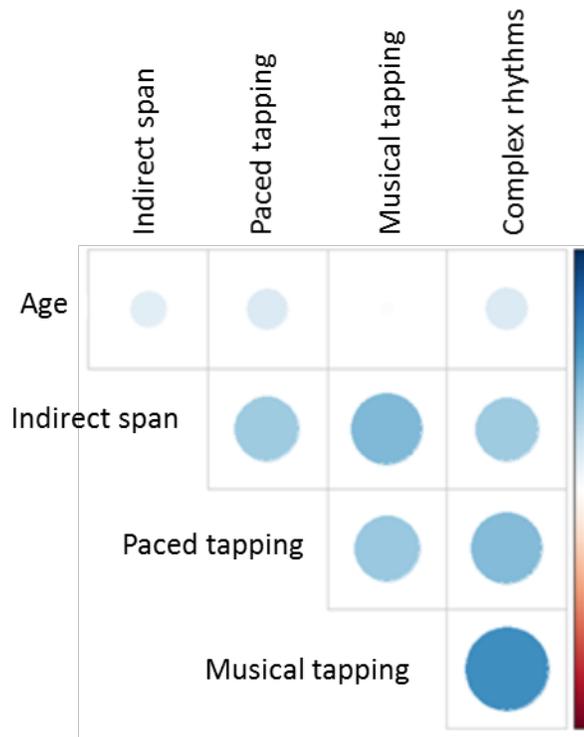


Figure X.

The second correlational analysis was computed on the same variables including sentences repetition scores but only on data of children with cochlear implants. Results showed a link between sentence repetition and complex rhythm performances ($r = .5$; CI [.17; .73], indirect span ($r = .45$; CI [.09; .73] and age ($r = .42$; CI [.03; .73]). Based on the results of this correlation analysis, we used a linear model to assess the best predictor of the sentences repetition score in children with CI. More precisely, we used a multiple linear regression model to test whether performance on sentence repetition could be better predicted by performance on complex rhythm, age or their interaction, controlling for differences in working memory (covariate). Results showed that complex rhythm performance best predicts performance in sentence repetition ($\beta = 11.49$, $SE = 4.67$, $t = 2.45$, $p < .05$).

Discussion

In this study we investigated several rhythmic abilities in children with NH and congenitally deaf children with CI. Different tasks assessed sensori-motor synchronization using stimuli of different acoustic complexity as well as production of more complex rhythmic sequences. We have also measured working memory in both populations and speech perception in children with CI. Our first aim was to assess potential deficits in temporal processing disentangling the role of basic auditory and cognitive processing. Our second aim was to investigate a possible relation of temporal processing deficits and speech perception. We will first discuss the synchronization deficit found in cochlear implanted children compared to their normal-hearing peers. We will then discuss the difficulty encountered by children with CI in realizing hierarchical temporal processing of auditory information and the links between this rhythmic competence and speech perception.

Cochlear-implanted congenitally deaf children have poor sensory-motor synchronization abilities

It is well known that postlingually deaf adults listening to music can better appreciate the rhythmic than the melodic structure (Looi, McDermott, McKay, & Hickson, 2008; McDermott, 2004). CI adults perform similarly to NH adults in rhythmic perception tasks such as tempo discrimination and discrimination or identification of rhythmic patterns; this is not the case in tone or melody discrimination tasks (Collins, Wakefield, & Feinman, 1994; Gfeller et al., 1997; Kong, Cruz, Jones, & Zeng, 2004). Unlike pitch perception, rhythm perception does not require a fine spectral analysis. Considering the technical limitations of the cochlear implant in terms of spectral processing and the rather good performance in terms of temporal processing of the auditory signal (Limb & Roy, 2014) (see Limb & Roy, 2014), the different performance of CI adults in melodic and rhythmic tasks can be interpreted in terms of the acoustic nature of the stimuli to be processed.

Phillips-Silver et al. (2015) used stimuli with different acoustic complexity to investigate the sensory-motor synchronization abilities of postlingually deaf adults wearing a CI. More precisely the authors used three versions of the same song. An original version, a version played on the piano, containing the rhythmic structure associated with the variations of pitch but reducing the spectral complexity, and a version played on percussions containing only the rhythmic structure. The authors then compared the tapping performances obtained with these stimuli to those obtained with visual or auditory metronomes. Results show better scores for NH than CI adults with all types of stimuli except the visual metronome. Moreover, while CI adults poorly synchronize to the original or piano versions of the music excerpt, their performance with the percussive version is similar to the one measured

with the auditory or visual metronomes (Phillips-Silver et al., 2015). Thus, for stimuli with a limited spectral complexity, postlingually deaf adults with CI can perceive rhythmic patterns, extract the beat, and synchronize to it.

The present results with congenitally deaf children with CI confirm what previously described with adults: CI children fail to achieve results close to those of NH children when tapping to a metronome even though the sound of the metronome is acoustically simple and clearly perceived. This result suggests that early auditory deprivation may have altered the abilities involved in sensorimotor synchronization tasks. According to the Adaptation and Anticipation model (van der Steen & Keller, 2013)(van der Steen et al., 2013), a tapping task would rely on two types of mechanisms. Anticipation that would allow predicting the appearance of future events and thus preparing the movement sufficiently in advance. Adaptation that would then allow correcting the time difference between the occurrence of the movement and that of the sensory entry. Thus, one possible interpretation of our results is to consider that, unlike adults who have become deaf, congenitally deaf children have more difficulties than children with NH to accurately anticipate the course of regular events and to adapt their motor response according to their errors even when the stimulus does not require a complex encoding of acoustic information. Interestingly, this deficit could have consequences on the quality of speech perception. Indeed, efficient speech perception requires temporal processing at different time scales: at phonemic level, the syllabic level, for word segmentation and at a suprasegmental/prosodic level.

Tierney & Kraus (2014) suggest that a precise auditory timing hypothesis (PATH) may be at the heart of enhanced phonological abilities in musicians via emphasized stimulus to brain entrainment (Tierney & Kraus, 2014). Deafness being the other side of the coin of musical expertise, it is possible that despite early cochlear implantation, sensory deprivation may affect the quality of temporal coding of acoustic stimuli. This may not be confined to the subcortical level but may extend to a broader network level, including the colliculo-cerebellar connections, the fronto-striatal motor loop and the dorsal pathway (Kral, Kronenberger, Pisoni, & O'Donoghue, 2016; Young et al., 2018). In other words, the audio-motor network involved in anticipating stimuli and adapting motor responses. Interestingly, the deficit observable in temporal processing of simple acoustic sequences becomes even more marked in rhythmic tasks requiring similar acoustic processing but more complex processing of temporal information.

Cochlear-implanted congenitally deaf children have a deficit in perceiving the hierarchical organization of auditory stimuli

While children with CI perform more poorly in the sensory-motor synchronization to the auditory metronome compared to the children with NH, the difference across groups is even more evident when they have to synchronize to a music excerpt and when they have to reproduce rhythmic sequences. This greater difficulty in tapping to the musical excerpts could be attributable, as in adults with CI (Phillips-Silver et al., 2015), to the difficulty in processing and extracting the beat of a stimulus that requires a fine spectral processing, highly challenging for a cochlear implant.

However, since this deficit is also present in tapping to a simple stimulus (auditory metronome), it cannot be fully explained in terms of the poverty of the spectral resolution of the implant. Moreover, complex rhythmic sequences that did not require any complex acoustic processing involved a more complex temporal structure than the isochronous metronomic sequence. Thus, it seems that a poor spectral representation cannot fully account for our results. Rather, there seems to be a truly temporal processing deficit. This is present already with the simplest stimulus (metronome) and is amplified with rhythmic sequences that require to build a more complex temporal structure such as tapping to musical excerpts and reproducing complex rhythms. To our knowledge this is the first time that this deficit is clearly isolated in children with CI.

Previous studies on children with CI testing rhythmic discrimination abilities presented contradictory results. Innes-Brown et al. (2013) tested the musical perceptual abilities of normal-hearing children (N = 8), implanted children (N = 6), and children with hearing aids (N = 5). Children with CI scored lower than normal-hearing children in pitch discrimination and timbre recognition tasks, but showed a similar performance in rhythmic discrimination. Interestingly, while the degree of deafness was a good predictor of performance in the pitch and timbre tests, it did not predict rhythmic performance. By contrast Stabej et al. (2012) tested about forty children with CI as well as a control group and found that children with CI are impaired in the discrimination of rhythmic patterns (Stabej et al., 2012). The results of this latter study on a rather large population go along with our findings of a deficit in temporal processing in children with CI. While it is true that the rhythmic reproduction task that we have used is quite challenging, one should note that children with NH, that were age matched to children with CI, were proficient in this task. Thus, children with CI have actually been challenged by the specific type of processing required in this type of task.

In considering the specificity of the processing deficit it is important to carefully examine the possible role of working memory in this type of task that requires to reproduce a rather long rhythmic sequence. All the more so that we replicated previous studies (Beer, Kronenberger, & Pisoni, 2011; Pisoni & Cleary, 2003) showing a working memory deficit in children with CI children. Firstly, it is important to recall that in this this task children tap the rhythmic sequence along with the tablet. Every rhythm is repeated 8 times and performance is attributed solely on the basis of the 8th repetition. Reproducing several times the same rhythm and tapping along to the rhythm are both supposed to reduce the working memory load. Moreover, we also used statistical modeling to control for possible effects of working memory differences. When adding working memory as a covariate to the linear regression, the performance in the rhythmic reproduction task remains significantly different in children with CI compared to children with NH. Thus, neither perceptual nor the memory skills of children with CI seem to be responsible for their deficit in this task.

The deficit in rhythmic discrimination (Stabej et al., 2012), the lack of correlation between rhythmic discrimination and degree of deafness (Innès-Brown, 2013) as well as our findings showing synchronization difficulties on acoustically simple but structurally complex stimuli lead us to hypothesize that congenitally deaf children with CI may have difficulty in perceiving a temporal hierarchical organization of sounds. This may have affect their ability to perceive language. This is supported by the finding that performance in rhythm reproduction, which is also the task wherein children with CI mostly differ from children with NH, best predicts performance in sentence repetition. This holds true when controlling for working memory variability.

The ability to organize sequential auditory events according to a hierarchical structure would indeed be of utmost importance in auditory (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Synchronization with music and complex rhythmic sequences requires two interrelated temporal skills: extracting the beat, the regular pulse, and organizing the beats into a hierarchical structure with different levels of prominence. In the context of language, this type of high-level cognitive operation would be relevant in structuring speech sounds as more or less prominent perceptual units. These perceptual units, called p-centers, are grouped according to accentuated patterns (i.e. alternation of weak-strong syllables). In the field of prosody, this organization is called metric phonology and makes it possible, for example, to perceive the syllables composing a statement with different levels of prominence (accent) and to apprehend the syllables of the statement boundaries as being the mostly emphasized in the hierarchy. More generally, different temporal constraints may

apply to the different levels of the entire hierarchy of speech from the prosodic to the phonetic level and may be affected by the ability of the system to temporally structure auditory information.

To conclude, it may be that congenitally deaf children with CI, unlike post-lingually deaf CI adults, suffer from a deficit of anticipation and/or temporal adaptation that is present with simple stimuli and accentuated with stimuli of increasing acoustic and temporal complexity. As suggested by Gfeller (1997) for post-lingually deaf CI adults, this deficit may be accompanied by a higher level temporal deficit related to the structuring of acoustic events in a hierarchical manner (Gfeller et al., 1997). Given the link found between language and rhythm deficit in children with CI, these conclusions suggest, as recommended by studies conducted in other clinical populations with speech perception deficit (Flaugnacco et al., 2015; Przybylski et al., 2013), that musical training, and more specifically rhythmic training, could have beneficial effects on the ability to perceive and produce speech in children with CI.

Baldassari, C. M., Schmidt, C., Schubert, C. M., Srinivasan, P., Dodson, K. M., & Sismanis, A. (2009).

Receptive language outcomes in children after cochlear implantation. *Otolaryngology - Head and Neck Surgery*, *140*(1), 114–119. <http://doi.org/10.1016/j.otohns.2008.09.008>

Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally Regular Musical

Primes Facilitate Subsequent Syntax Processing in Children with Specific Language Impairment. *Frontiers in Neuroscience*, *10*, 245. <http://doi.org/10.3389/fnins.2016.00245>

Beer, J., Kronenberger, W. G., & Pisoni, D. B. (2011). Executive function in everyday life: implications

for young cochlear implant users. *Cochlear Implants International*, *12 Suppl 1*(Suppl 1), S89-91. <http://doi.org/10.1179/146701011X13001035752570>

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on

speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, *51*(4), 760–771. <http://doi.org/10.1016/j.neuropsychologia.2013.01.006>

Caselli, M. C., Rinaldi, P., Varuzza, C., Giuliani, A., & Burdo, S. (2012). Cochlear Implant in the Second

Year of Life: Lexical and Grammatical Outcomes. *Journal of Speech Language and Hearing Research*, *55*(2), 382. [http://doi.org/10.1044/1092-4388\(2011/10-0248\)](http://doi.org/10.1044/1092-4388(2011/10-0248))

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech

production abilities: Evidence from prelingually deaf children. *Neuropsychology*, *29*(1), 102–107. <http://doi.org/10.1037/neu0000115>

- Castellanos, I., Kronenberger, W. G., Beer, J., Colson, B. G., Henning, S. C., Ditmars, A., & Pisoni, D. B. (2015). Concept formation skills in long-term cochlear implant users. *Journal of Deaf Studies and Deaf Education, 20*(1), 27–40. <http://doi.org/10.1093/deafed/enu039>
- Collins, L. M., Wakefield, G. H., & Feinman, G. R. (1994). Temporal pattern discrimination and speech recognition under electrical stimulation, *96*(November), 2731–2737.
- Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science, 14*(1), 69–82. <http://doi.org/10.1111/j.1467-7687.2010.00960.x>
- Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences, 1316*(1), 87–104. <http://doi.org/10.1111/nyas.12406>
- Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music Training Increases Phonological Awareness and Reading Skills in Developmental Dyslexia: A Randomized Control Trial. *Plos One, 10*(9), e0138715. <http://doi.org/10.1371/journal.pone.0138715>
- Fowler, C. A. (1979). “Perceptual centers” in speech production and perception. *Perception & Psychophysics, 25*(5), 375–388. <http://doi.org/10.3758/BF03199846>
- Geers, A. E., Moog, J. S., Biedenstein, J., Brenner, C., & Hayes, H. (2009). Spoken language scores of children using cochlear implants compared to hearing age-mates at school entry. *Journal of Deaf Studies and Deaf Education, 14*(3), 371–385. <http://doi.org/10.1093/deafed/enn046>
- Geers, A. E., & Sedey, A. L. (2011). Language and Verbal Reasoning Skills in Adolescents With 10 or More Years of Cochlear Implant Experience. *Ear & Hearing, 32*, 39–48. <http://doi.org/10.1097/AUD.0b013e3181fa41dc>
- Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., & Knutson, J. F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and hearing, 18*(3), 252-260.
- Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews, 86*(December), 150–165. <http://doi.org/10.1016/j.neubiorev.2017.12.002>

- Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic training on children with hearing loss. *Hearing Research*, 351(May), 11–18.
<http://doi.org/10.1016/j.heares.2017.05.006>
- Holt, C. M., Yuen, I., & Demuth, K. (2017). Discourse Strategies and the Production of Prosody by Prelingually Deaf Adolescent Cochlear Implant Users. *Ear and Hearing*, 38(2), e101–e108.
<http://doi.org/10.1097/AUD.0000000000000362>
- Jiam, N. T., Caldwell, M., Deroche, M. L., Chatterjee, M., & Limb, C. J. (2017). Voice Emotion Perception and Production in Cochlear Implant Users. *Hearing Research*.
<http://doi.org/10.1016/j.heares.2017.01.006>
- Jones, M. R. (2016). Musical time. In *The Oxford handbook of music psychology, 2nd ed.* (pp. 125–141). New York, NY, US: Oxford University Press.
- Jun, S.-A., & Fougeron, C. (2002). *Realizations of accentual phrase in French intonation. Probus* (Vol. 14). Retrieved from
<https://pdfs.semanticscholar.org/7c75/1788787524ca789c0b53a32ef42763e45e58.pdf>
- Kong, Y.-Y., Cruz, R., Jones, J. A., & Zeng, F.-G. (2004). Music Perception with Temporal Cues in Acoustic and Electric Hearing. *Ear and Hearing*, 25(2), 173–185.
<http://doi.org/10.1097/01.AUD.0000120365.97792.2F>
- Kral, A., Kronenberger, W. G., Pisoni, D. B., & O’Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: a connectome model. *The Lancet Neurology*, 15(6), 610–621. [http://doi.org/10.1016/S1474-4422\(16\)00034-X](http://doi.org/10.1016/S1474-4422(16)00034-X)
- Lederberg, A. R., Schick, B., & Spencer, P. E. (2013). Language and literacy development of deaf and hard-of-hearing children: Successes and challenges. *Developmental Psychology*, 49(1), 15–30.
<http://doi.org/10.1037/a0029558>
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5(3), 253–263.
- Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, 308, 13–26.
<http://doi.org/10.1016/J.HEARES.2013.04.009>
- London, J. (2004). *Hearing in Time. Musicae Scientiae* (Vol. 10). Oxford University Press.

<http://doi.org/10.1093/acprof:oso/9780195160819.001.0001>

Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing, 29*(3), 421–434.

<http://doi.org/10.1097/AUD.0b013e31816a0d0b>

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification, 8*(2), 49–82. <http://doi.org/10.1177/108471380400800203>

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Turgeon, C., Lepore, F., & Peretz, I. (2015). Cochlear implant users move in time to the beat of drum music. *Hearing Research, 321*, 25–34.

<http://doi.org/10.1016/j.heares.2014.12.007>

Pisoni, D. B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*(1 Suppl), 106S–20S.

<http://doi.org/10.1097/01.AUD.0000051692.05140.8E>

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., ... Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology, 27*(1), 121–31. <http://doi.org/10.1037/a0031277>

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects.

Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 336(1278), 367–73. <http://doi.org/10.1098/rstb.1992.0070>

Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., & Vatovec, J. (2012). The music perception abilities of prelingually deaf children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 76*(10), 1392–400. <http://doi.org/10.1016/j.ijporl.2012.07.004>

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience, 8*(November), 1–9.

<http://doi.org/10.3389/fnhum.2014.00949>

Tierney, A., & Kraus, N. (2015). Evidence for multiple rhythmic skills. *PLoS ONE, 10*(9), 1–14.

<http://doi.org/10.1371/journal.pone.0136645>

Torppa, R., Faulkner, A., Huotilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014).

The perception of prosody and associated auditory cues in early-implanted children: the role of

auditory working memory and musical activities. *International Journal of Audiology*, 53(3), 182–91. <http://doi.org/10.3109/14992027.2013.872302>

Truy, E., Ve Lina-Granade, G., Jonas, A.-M., Ve Martinon, G., Maison, S., Girard, J., ... Morgon, A. (1998). *Comprehension of language in congenitally deaf children with and without cochlear implants. International Journal of Pediatric Otorhinolaryngology* (Vol. 45). Retrieved from https://ac-els-cdn-com.lama.univ-amu.fr/S0165587698000974/1-s2.0-S0165587698000974-main.pdf?_tid=41326215-0a7c-4cec-b81a-1feb429903ca&acdnat=1541350400_7e719482c53bc130abec87c58305220b

van der Steen, M. C. M., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7(June), 253. <http://doi.org/10.3389/fnhum.2013.00253>

Wang, Y., Shafto, C. L., & Houston, D. M. (2018). Attention to speech and spoken language development in deaf children with cochlear implants: a 10-year longitudinal study. *Developmental Science*, (March), e12677. <http://doi.org/10.1111/desc.12677>

Wechsler-Kashi, D., Schwartz, R. G., & Cleary, M. (2014). Picture Naming and Verbal Fluency in Children With Cochlear Implants. *Journal of Speech Language and Hearing Research*, 57(5), 1870. http://doi.org/10.1044/2014_JSLHR-L-13-0321

Young, N., Feng, G., Ingvalson, E. M., Grieco-calub, T. M., Roberts, M. Y., & Ryan, M. E. (2018). Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients, (January). <http://doi.org/10.1073/pnas.1717603115>

Expérience n°2



Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Speak on time! Effects of a musical rhythmic training on children with hearing loss

Céline Hidalgo ^{a, d, *}, Simone Falk ^{a, b, c}, Daniele Schön ^d^a Aix Marseille Univ, CNRS, LPL, Aix-en Provence, France^b Clinical & German Linguistics, Ludwig-Maximilians-University Munich, Germany^c Laboratoire Phonétique et Phonologie, UMR 7018, CNRS, Université Sorbonne Nouvelle Paris-3, France^d Aix Marseille Univ, INSERM, INS, Inst Neurosci Syst, Marseille, France

ARTICLE INFO

Article history:

Received 19 October 2016

Received in revised form

5 May 2017

Accepted 9 May 2017

Available online xxx

Keywords:

Hearing loss

Children

Temporal accommodation

Speech production

Interaction

Rhythmic training

ABSTRACT

This study investigates temporal adaptation in speech interaction in children with normal hearing and in children with cochlear implants (CIs) and/or hearing aids (HAs). We also address the question of whether musical rhythmic training can improve these skills in children with hearing loss (HL). Children named pictures presented on the screen in alternation with a virtual partner. Alternation rate (fast or slow) and the temporal predictability (match vs mismatch of stress occurrences) were manipulated. One group of children with normal hearing (NH) and one with HL were tested. The latter group was tested twice: once after 30 min of speech therapy and once after 30 min of musical rhythmic training. Both groups of children (NH and with HL) can adjust their speech production to the rate of alternation of the virtual partner. Moreover, while children with normal hearing benefit from the temporal regularity of stress occurrences, children with HL become sensitive to this manipulation only after rhythmic training. Rhythmic training may help children with HL to structure the temporal flow of their verbal interactions.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Conversation requires that speakers accommodate to the verbal behavior of their conversational partner. This accommodation can alter the content (e.g. semantic system, [Garrod and Anderson, 1987](#)), the form of the exchange (e.g. intensity, [Natale, 1975](#)) or temporal aspects of the conversation (e.g., turn-taking, speaking rate, [Beñuš et al., 2011](#); [Street, 1984](#)). Generally, accommodation can lead to greater similarity of interlocutors' verbal patterns (e.g., [Abney et al., 2014](#)). It also plays an important role in enhancing mutual comprehension ([Garrod and Pickering, 2004](#)) and cooperation ([Manson et al., 2013](#)), and in reducing the social distance between interlocutors ([Giles et al., 1991](#)). In particular *temporal* accommodation between interlocutors, that is mutual adaptation

of temporal structure (speech rate, turn-timing), is associated with positive perception (pleasantness) of a conversation ([Warner et al., 1987](#)). This type of accommodation evolves from very early on, in the interactions between infants (4 months) and their mothers ([Jaffe et al., 2001](#)). Importantly, temporal accommodation implies the capacity of interlocutors to *anticipate* various temporal characteristics of the partner's speech, such as the time of important discourse units, ends of utterances or turns ([Garrod and Pickering, 2015](#)). Temporal predictions aid interlocutors to "be on time" with their own verbal production, warranting a fluent verbal exchange and sustaining the flow of interaction.

Anticipatory processes involved in the development of conversational skills were already found in children as young as three years using eye tracking ([Casillas and Frank, 2013](#)). When watching a dialogue, three-year olds shifted their gaze onto the next speaker although the current speaker had not yet completed his turn. These results indirectly show that children develop early anticipatory abilities inherent to the conversation.

[Hilbrink et al. \(2015\)](#) obtained a more direct measure of these abilities by analyzing dialogues between mothers and children aged from 3 to 18 months. Their analysis revealed that temporal patterns of speech turns of mother-infant conversations (5 months) were similar to those found in adults (i.e., with little overlap of

Abbreviations: NH, normal hearing; HL, hearing loss; P-center, perceptual center; IWI, inter-word-interval; ITI, inter-turn-interval; CI, cochlear implant; HA, hearing aid

* Corresponding author. Institut de Neurosciences des Systèmes, UMR 1106, INSERM, Aix-Marseille University, 27 boulevard Jean Moulin, faculté de Médecine de la Timone, 13005 Marseille, France; Laboratoire Parole et Langage, UMR 7309, CNRS, Aix-Marseille University, 5 Avenue Pasteur, 13100 Aix-en-Provence, France.

E-mail address: celine.hidalgo@univ-amu.fr (C. Hidalgo).

<http://dx.doi.org/10.1016/j.heares.2017.05.006>

0378-5955/© 2017 Elsevier B.V. All rights reserved.

vocalizations and a small time interval between two turns). However, temporal accommodation and development of anticipatory skills in the production of children at verbal age await further clarification.

In clinical populations, it has been shown that using rhythmic cues and training can enhance temporal predictions in speech and thereby, may aid temporal accommodation in conversation. For example, patients with dysarthria (Liss et al., 2009) display irregular vocalic and consonantal segment durations altering speech rate and impacting conversational abilities. However, regular stress distributions in the speech of an interlocutor, allowing for enhanced temporal predictions, can help dysarthric patients to better accommodate to the speech of their conversational partner (Späth et al., 2016). Other clinical studies have highlighted that verbal and non-verbal rhythmic training, in particular by combining auditory and motor stimulation, can have positive effects in patients suffering from language disorders involving temporal deficits. For example, patients with non-fluent aphasia, improved their speech production (i.e. articulatory quality) when speech was accompanied by a regular pulse (Stahl et al., 2011). Rhythmic auditory cues accompanying speech can also enhance fluency in children and adults who stutter (Andrews et al., 1982). Children with SLI and dyslexia improved their grammatical judgments after listening to a regular prime (compared to an irregular prime, Przybylski et al., 2013). Children with dyslexia improved their phonological and reading abilities after an active musical training and this improvement in speech abilities was correlated with an improvement in their rhythmic abilities (Flaugnacco et al., 2015). Children with non-verbal autism spectrum disorders increased the range and the complexity of their vocal productions, after using an audio-motor training (15 min, 5 times a week during 8 weeks) as an intervention program (Wan et al., 2011). These examples show that listening to regular rhythms or training rhythmic skills helps patients who suffer from deficits in temporal processing to better encode and predict the temporal structure of speech.

A particularly vulnerable population, with regard to the role of temporal predictions for their communicative skills, are children with hearing loss (HL). Children with cochlear implants (CIs) and/or hearing aids (HAs) show high variability in spoken language achievement and difficulties in communicational skills compared to children with normal hearing (NH) (van Wieringen and Wouters, 2015). Pragmatic skills develop relatively late in children with hearing loss using both speech and sign language, compared to children with NH (Bebko et al., 2003). Studies in children with HL describe inappropriate use of pragmatic features in speech turns (Most et al., 2010), greater presence of silences, communication breakdowns (Tye-Murray, 2003) and excessively long speech turns (Toe and Paatsch, 2013), eventually reducing the smoothness and balance of verbal exchanges.

These studies on pragmatic skills in children with HL clearly suggest conversational difficulties but miss a link with more general temporal processing deficits. Interestingly, the few studies addressing temporal prediction and processing in children with HL have been carried out in the music domain. While children with CIs or HAs are able to discriminate rhythmic patterns (Innes-Brown et al., 2013) children with CIs are less precise than children with NH (Roy et al., 2014; Stabej et al., 2012). Electrophysiological studies using the mismatch negativity protocol (MMN) show that children with CIs are less sensitive to temporal manipulation of a regular auditory structure compared to NH children (Petersen et al., 2015; Torppa et al., 2012). Interestingly, it was recently shown that, in a sentence repetition task, musical rhythmic priming of speech, by engaging temporal predictions, facilitated phoneme and word perception and production abilities of children with CIs and HAs (Cason et al., 2015b).

In light of these findings, we wanted to test whether training temporal prediction capacities in children with HL via musical rhythm may impact their communication and accommodation skills in production. More precisely, the aims of the present study were twofold. First, focusing on children with NH, we investigated how rhythmic predictability of turn-taking impact on temporal accommodation skills at 5–6 years of age. Second, from a more clinical standpoint, we examined the hypothesis that rhythmic training influences these skills in French speaking children with HL wearing CIs and/or HAs. Indeed, CI technology can deliver temporal information better than spectral cues and several studies have shown a good performance of CI children in rhythmical musical tasks (e.g. Innes-Brown et al., 2013; Roy et al., 2014).

Inspired by adaptation studies with adults (Himberg et al., 2015; Kawasaki et al., 2013; Späth et al., 2016), we developed a child-friendly, novel procedure of a turn-taking paradigm (an “alternating picture naming task”) manipulating the difficulty of temporal anticipation required to perform the task. In a first Experiment, children with NH named pictures on the screen in alternation with a virtual partner. The temporal predictability of stress occurrences¹ and thus of the turns was manipulated as well as the alternation rate. We measured children's temporal adaptation and prediction skills by analysing the timing of the stressed syllables of words with respect to those of the virtual partner. We hypothesized that children should adopt a regular rhythm in their own turn-production (i.e., producing their words consistently with respect to the virtual partner, Himberg et al., 2015) depending on the alternation rate of the partner. Moreover, we hypothesized that altering the number of syllables per word and thus, lowering the predictability of time of stress occurrences should render the task more difficult and deteriorate children's consistency in turn-taking. In a second Experiment, we used the same procedure to measure the impact of rhythmic training on speech production and more precisely on the temporal prediction abilities in speech in children with HL. Based on previous findings with other language impaired populations (Flaugnacco et al., 2015; Przybylski et al., 2013), we expected children with HL to benefit from rhythmic training in their adaptation skills, especially in the less predictable context.

2. Experiment 1

2.1. Material and methods

2.1.1. Participants

Sixteen children with NH were recruited from a kindergarten in Marseille. They were aged from 5 to 6 years old (mean age = 65 months, $SD = 5$ months) and were all native-speakers of French without any visual, speech, cognitive or hearing disorder. This experiment has been approved by the ethics committee Sud Méditerranée I (n°ID RCB: 2015-A01490-49). All children's parents signed a consent form.

2.1.2. Stimuli

One hundred sixty pictures from 16 taxonomic categories were selected from the French BD2I database (Cannard et al., 2005). Half of them depicted monosyllabic words, half were disyllabic. The selected pictures were named correctly by over 90% of a normative sample of 3–8 year old children (monosyllabic words: mean score = 93.9%, $SD = 5.3\%$; disyllabic words: mean score = 98.8%, $SD = 1.6\%$).

¹ We use the term “stress” throughout the manuscript in its broad sense (accent, emphasis), and not in terms of “lexical stress” which is not a feature of French, the native language of the participants in the present study.

The picture names were recorded in a soundproof booth by a female native French speaker (henceforth, the virtual partner). An additional set of three words was recorded by an 8-year-old boy in order to construct a model for the alternated naming procedure. The perceptual centers of the syllables (*p-centers*) were estimated at 2/3 of the amplitude rise in the sound envelope using a semi automatic procedure (Cummins and Port, 1998). These estimated *p-centers*, described as being rhythmic anchor-points in speech, were used as points of reference for the adjustment of the interval between the words uttered by the adult virtual partner (i.e., the Inter-Turn-Interval, ITI). In disyllabic words, the *p-center* on the stressed syllable (i.e., the final syllable of each French accentual phrase) was chosen.

2.1.3. Procedure

Before the beginning of the Experiment, children were familiarized with the pictures. To this end, all the pictures used in the task were presented on a sheet and named by the participant to ensure the good identification and naming of pictures as well as the articulatory correctness of the words. After this familiarization phase, the participants were seated in front of a computer screen and instructed to name the pictures, displayed on the screen, in alternation with a virtual partner. On the screen, the child saw the pictures of the virtual partner (left-hand side) and his/her own pictures (right-hand side, see Fig. 1) and was asked to name the picture after the virtual partner.

In order to manipulate temporal predictability of the turns, pairs of pictures were created by varying their syllable number (“word regularity”) which modified the timing of words and their stress occurrences. In a highly predictable condition, the pair was *matched*

in the number of syllables (i.e. the names of both pictures had the same number of syllables, mono- or disyllabic, respectively). In the less predictable condition, the pair was *mismatched* (i.e. one picture name was a monosyllabic word and the other a disyllabic word). Four blocks were built containing each 10 pairs of pictures in the *matched* condition (1 with monosyllabic words and 1 with disyllabic words) and 4 blocks of 10 pairs of pictures in the *mismatched* condition. Pictures in a given pair and pair order were carefully controlled in order to avoid semantic, categorical or phonological priming between words of the same pair or between two successive pairs. In a given block, the alternation rate was determined by the ITI of the virtual partner. It was either a fast rate (i.e., ITI of 2600 ms between the *p-centers* of two successive words of the virtual partner, see Fig. 1) or slow (i.e., ITI of 3200 ms). This resulted in a 2 by 2 factorial design manipulating word regularity (matched or mismatched) and alternation rate (slow or fast) across blocks. The order of block presentation was counterbalanced across participants.

Each block started with 3 pairs of pictures that were named by the adult virtual partner in perfectly timed alternation with the voice of the 8-year-old boy in order to establish the rate of alternation in an implicit manner. Next, 10 pairs of pictures were presented successively at the bottom of the screen in a scrolling manner and the child was instructed to name the pictures continuing the alternation. Children naturally tended to name the images in between the two successive pronunciations of the virtual partner. No information concerning when to pronounce the name of the image nor on word regularity and alternation rate were given to the child. A block was not interrupted even if the child misnamed or did not name a picture. The child was praised and encouraged to

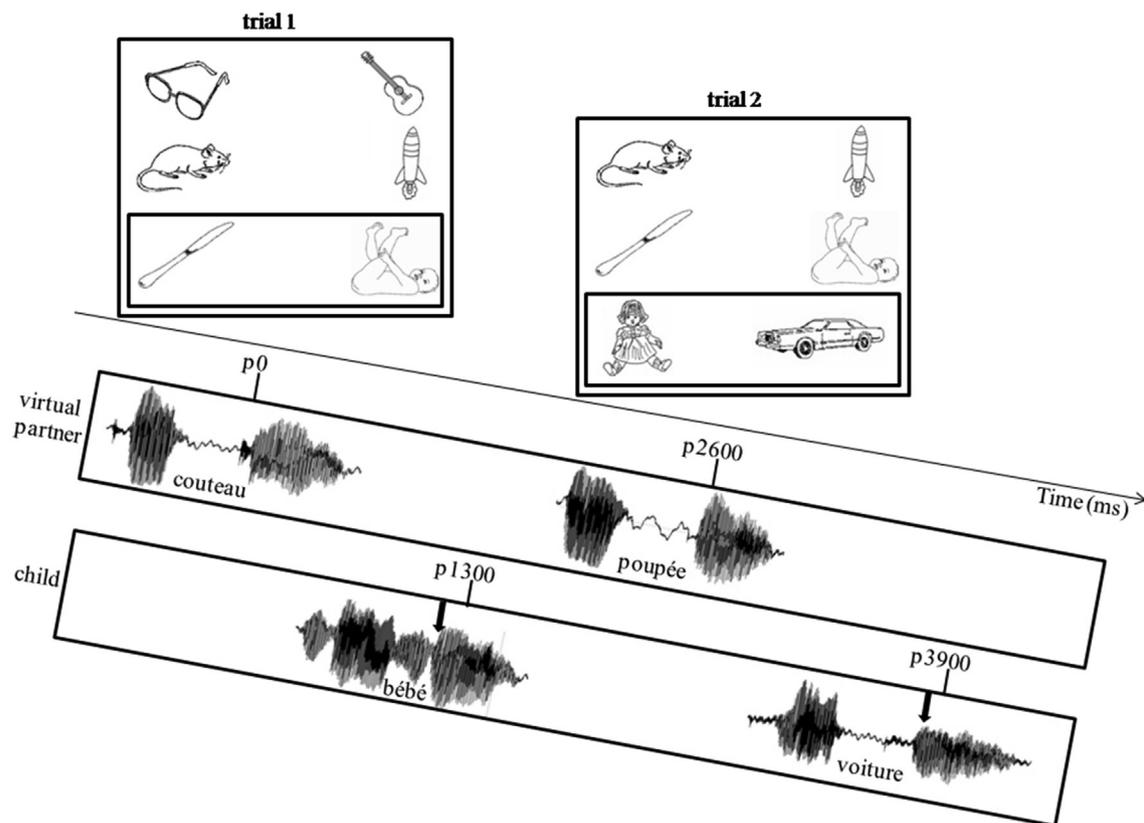


Fig. 1. Representation of the time course of 2 trials of a fast alternation block. The thick rectangle indicates the current pair of pictures. The virtual partner names the pictures on the left (trial 1: knife fr. *couteau*), while the child has to name the pictures on the right (trial 1: baby, fr. *bébé*). Corresponding representation of the speech signal. The ticks indicate the location of the virtual partner's *p-centers* (p 0, p 2600) and, for the child, the expected location of the child's *p-centers* (p 1300, p 3900). The actual child's *p-centers*, in this example, are indicated by an arrow and are slightly ahead the expected moment.

continue after completion of each block.

Stimuli delivery was programmed using the software *Presentation* (Neurobehavioral Systems 18.0). Pictures were displayed on a laptop screen (DELL Latitude E5530, screen resolution 1366 × 768). The volume for the voice of the virtual partner, delivered by a CREATIVE sound card (X-Fi 5.1) and two Sony loudspeakers (SRS A205), was adjusted for each child at a comfortable level. The naming performance of the participant was recorded with a head-mounted microphone (Sennheiser HSP) linked to a ZOOM H4n numeric recorder at a 44 Hz sampling rate.

2.1.4. Analysis

Inter-word-intervals between the child and the virtual partner and the relative timing of children's words to the words of the virtual partner were chosen as measures to evaluate temporal accommodation (Himberg et al., 2015). In order to determine the Inter-Word-Interval (IWI) between the virtual partner and the child, we estimated the distance of the p-centers of the stressed syllables within each pair of words as pronounced by the virtual partner and the child (see Fig. 1). The IWI was chosen in order to evaluate children's accommodation to the different alternation rates (fast and slow). Intervals were excluded from analysis if the child made a hesitation (e.g., *euuh*), erroneously placed an article before naming the object or spoke simultaneously with the virtual partner. This resulted in rejecting approximately 10% of word-intervals. Then, we analyzed the timing of children's stressed syllables in relation to those of the virtual partner. Because of the highly predictable rhythm of alternation, it was expected that children would anticipate the time of the words (their stressed syllable) of the virtual partner and, consequently, produce their own words (stressed syllables) in a stable temporal relationship to the virtual partner. This anticipatory behavior was confirmed in a pilot study on three children. A preferred and particularly stable temporal relation in rhythmic tasks requiring an action in response to a regular rhythmic auditory structure is the anti-phase (i.e., the mid-point between two periodically recurring auditory events; Repp, 2005; Volman and Geuze, 2000). Thus, the timing of the child's stressed syllables was analyzed, as a default, with respect to this 'expected' moment in time representative of a temporally perfectly regular alternation (i.e., half of the ITI of the virtual partner, notably, 1300 ms and 1600 ms after the p-center of the virtual partner for the slow and fast condition, respectively). The analysis was performed using circular statistics (CircStat Toolbox, Berens, 2009). Thereby, each child's response is expressed as a vector in a circular space with respect to the expected moment in time (see Supplemental material, appendix A for more details). In this space, the angle represents the asynchrony (i.e., the distance between the child's production and the expected moment in time) and the length of the vector represents the consistency (i.e., the variability of the child's asynchrony across trials). Consistency is particularly useful to estimate the child's temporal accommodation skills as more consistent (less variable) performances are indicative of a better temporal representation of the reciprocal turn-structure. Previous to any statistical inference, the Rayleigh test was used to verify that performances were not randomly distributed in the circular space (that is, not uniformly distributed). None of the children assessed in this study had a uniform distribution, showing that they had indeed produced the stressed syllables of the words in a consistent and anticipatory fashion with respect to the virtual partner's stresses.

2.2. Results

Three 2 × 2 repeated-measures analysis of variance (RM-ANOVA) with the factors Alternation rate (*fast* versus *slow*) and

Word regularity (*matched* versus *mismatched stress occurrences*) and the mean Inter-Word-Interval and consistency as the dependent variables were performed.

IWI analyses revealed a main effect of alternation rate: children with NH showed longer intervals in the slow than in the fast alternation condition ($F(1,15) = 24.99, p < 0.001$; see Fig. 2A left-hand side). Consistency analyses showed that children were less variable (i.e. more consistent) in the *fast* than in the *slow* condition (main effect of rate: $F(1,15) = 6.85, p = 0.019$). In addition, they showed a tendency to be less variable when the pairs of words had a regular instead of an irregular stress pattern (main effect of regularity: $F(1,15) = 4.09, p = 0.061$; see Fig. 2A right-hand side).

2.3. Discussion

These findings show that children with NH, as young as 5 years of age, are sensitive to temporal variations (alternation rate and the timing of words' stressed syllables) in the turns of their interlocutor. As a matter of fact, results on IWI revealed that children are able to adapt their naming performance to the rate of the virtual partner. This temporal adaptation (fast vs. slow: $\Delta = 157$ ms) is an important finding as it demonstrates that children were not simply reacting to the visual stimulus they had to name, but they were indeed taking into account changes in the temporal pace of alternation. Hence, our task, as intended, was apt to measure children's temporal attunement abilities in a pseudo-interactive setting.

We also observed a trend for facilitated adaptation when sequences of words were more predictable in their timing properties, that is in the *matched* condition, wherein the child's and the virtual partner's words had the same number of syllables (i.e. the same stress patterns). Indeed, the mismatched condition seemed to perturb the dynamics of the interaction by slightly degrading the phase-locking between the child and the virtual partner. Note that a disyllabic word with a final stress needs to be initiated earlier than a monosyllabic word. Because the alternating naming task requires the child to anticipate the moment of her/his turn in relation to the naming interval of the virtual partner, changing the number of syllables of words may degrade anticipatory strategies or put higher demands on the anticipatory mechanism in general.

3. Experiment 2

In Experiment 2, we used the same paradigm in order to investigate the capacities of temporal accommodation in children with CIs or HAs following a speech therapy session and a rhythmic musical training.

We hypothesized that children with HL would benefit more from a rhythmic training session in their adaptation skills, compared to a speech therapy session. This effect should be more evident in the mismatched condition, that is when the temporal predictability of word exchanges was more difficult.

3.1. Material and methods

3.1.1. Participants

Fifteen children with HL aged from 5 to 9 years (mean age = 89 months, $SD = 17$ months) were recruited from the Center d'Action Médicale Sociale Précoce in Marseille (children aged from 5 to 6 years) and the Service de Soutien à l'Éducation Familiale et à la Scolarisation in Marseille (children aged 6–9 years). Some of the children wore a unilateral CI, others wore bilateral CIs, some wore a CI in conjunction with a conventional HA on the contralateral ear and other HAs. Even if CI delivers an electric signal to the auditory nerve versus an amplified acoustic signal for HA, users of both devices may benefit from a rhythmic training. The children suffered

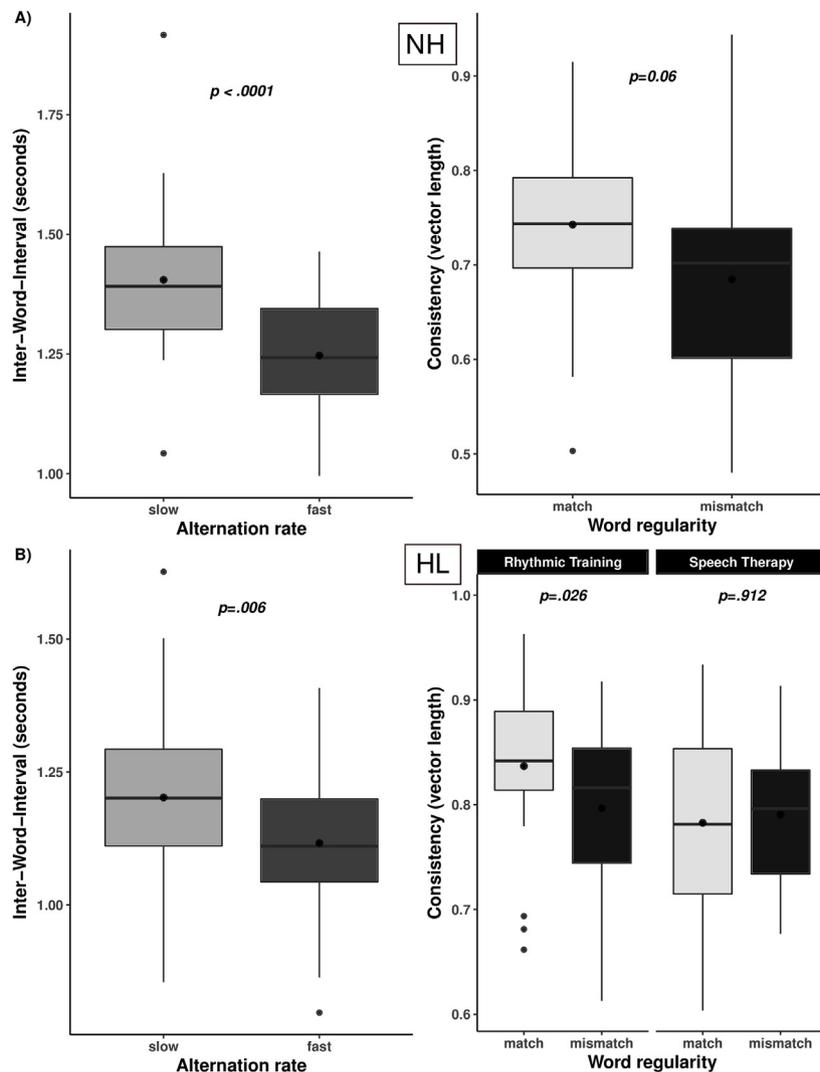


Fig. 2. Panel A: Children's (NH) temporal accommodation with a virtual partner. Error bars represent the standard error of the mean. Left: mean IWIs in seconds between the p-centers of the stressed syllables of the child and those of the virtual partner in the slow and fast conditions. Right: Consistency (vector length) of the child's word productions as a function of word regularity. Panel B: Temporal accommodation with a virtual partner in children with HL. Left: Mean IWIs in seconds between the p-centers of the child and those of the virtual partner in the slow and fast conditions. Right: Consistency (vector length) of the child's word production as a function of word regularity in both training conditions. P-values are results of post-hoc tests.

from different degrees of hearing loss (from profound to moderate, see Table 1 in the Supplemental material, [appendix B](#), where also the aided thresholds and other details of the participants are given). They were all native French speakers, educated in oral communication, attended mainstream primary school, and were free from known visual, speech or other cognitive impairment. Data of two of the 15 participants had to be excluded as they were unable to perform the task. This experiment has been approved by the ethics committee (n°ID RCB: 2015-A01490-49, Sud Méditerranée I). All children's parents signed a consent form.

3.1.2. Stimuli, procedure and analyses

The stimuli, the task and data analyses were the same as in Experiment 1. However, participants of Experiment 2 performed the naming task twice. Once, the measure was taken 30 min after a training session focusing on musical rhythmic stimulation which was meant to enhance temporal prediction skills in children. The training included several rhythmic audio-motor games (for more details, see Supplemental material, [appendix C](#)). As a comparison, a 'baseline' measure was taken 30 min after a speech therapy session

that focused on verbal stimulation in a classical speech therapy setting. Indeed, in order to avoid a general effect of training on arousal or mood that would render results difficult to interpret ([Thompson et al., 2001](#)), we opted for this 'baseline' measure instead of a non-stimulation baseline. Both types of stimulations lasted 30 min and used a playful and joyful setup. The 2 sessions were separated by one week on average and the order of stimulation was counterbalanced across children.

3.2. Results

Data were entered into a three $2 \times 2 \times 2$ RM ANOVA with Alternation rate (*fast* versus *slow*), Word regularity (*matched* versus *mismatched stress occurrences*), and Training type (*speech therapy* versus *rhythmic training*) as within subject factors and IWI (Inter-Word-Interval) and consistency (i.e. regularity) dependent variables.

Analyses on the IWI revealed a main effect of alternation rate on mean IWI: children with HL produced longer IWIs in the slow than in the fast condition ($F(1,12) = 10.95$, $p = 0.006$; see [Fig. 2B](#)

left-hand side). Consistency analyses showed that children with HL were generally more consistent in the *fast* than in the *slow* condition (consistency: $F(1,12) = 8.72, p = 0.012$).

As Fig. 2B (right-hand side) shows, we found no main effect of regularity on consistency ($p = 0.414$) but an interaction effect of training type and word regularity ($F(1,12) = 7.99, p = 0.015$).

3.3. Discussion

Children with HL were generally able to adapt their rate of production to the rate of the virtual partner and were generally more consistent in the *fast* than in the *slow* condition as NH children.

However, children with HL showed no differences in consistency when the pairs of words had a regular or an irregular number of syllables after the speech therapy session. Most importantly, after musical training, children with HL became more consistent in their alternation pattern in the matching than in the non-matching condition. This suggests that they were more sensitive to the temporal regularity of the speech exchanges after the rhythmic training. Note that (the younger) children with NH in Experiment 1 responded in a similar way (i.e., *matched* > *mismatched* stress occurrences), but without musical training.

4. General discussion

In sum, the present study showed that children with NH as young as 6 years of age were able to temporally anticipate and accommodate to the turn-structure of a virtual partner in a non-random way. Results showed that they were sensitive to the temporal predictability of turns and that they temporally adapted their speech production better to faster than slow turn rhythms. Children with HL, after having received speech or musical training also managed to perform the task and adapted better to fast than to slow turn rhythms. However, in contrast to the younger NH children, they did not benefit from more predictable turn-structure, unless they received musical rhythmic training. Most importantly, this is, to our knowledge, the first study to provide evidence that a rhythmic training of only 30 min may enhance the capacity of children with HL to accommodate to a temporally predictable flow of turn exchanges in an interactional setting. This effect points to potential benefits of rhythmic training on audio-motor anticipatory skills in children with HL. Rhythmic training may also have improved the extraction of regularities along the series of speech exchanges. By grouping discrete events (e.g. words) in a more hierarchical structure, children may have better perceived the metrical structure of the turns (i.e., the recurrence of stresses/syllables), allowing for a more precise temporal prediction of their own and their partner's turns.

Previous work has shown that musical training in adults enhances connectivity within a dorsal network comprising the auditory and the motor cortex (Chen et al., 2008). This network plays an important role in auditory predictions (Schubotz, 2007) and attentional dynamics (Schroeder et al., 2010). There is a possibility that rhythmic training in children with HL may stimulate the sensori-motor dorsal pathway and thereby improve children's abilities to adapt to a regular temporal structure (Vuust and Witek, 2014) or more generally improve their ability to perceive and understand rhythmical and metrical structure (Cason et al., 2015a). These different possibilities should be investigated in further detail in future studies.

Overall, our results are in line with previous studies showing that sensori-motor rhythmic training enhances auditory processing in different forms of language impairment at cortical and subcortical levels (Flaugnacco et al., 2015; Fujii and Wan, 2014; Tierney

and Kraus, 2013). Concerning the sensorimotor rhythmic capacities of adults with CIs, their lack of synchronization capacity on a complex musical stimulus (Phillips-Silver et al., 2015) could be indicative of the fact that meter (i.e., the regular recurrence of a beat or accents) perception and/or production is underdeveloped in hearing loss. Our data in children with HL seem to confirm these previous findings in adults. Interestingly, similarly to adults listening to complex auditory stimuli, children with HL, even after a speech therapy session, do not take advantage of the temporal regularity of stress occurrences in our task, while children with NH do. As metrical perception is shaped by auditory experience (Hannon and Trehub, 2005), musical practice may refine this type of perception in both music (Geiser et al., 2010), and speech (Marie et al., 2011). Developing the perception of meter in children with hearing loss may improve their general temporal anticipatory skills and thus enhance their ability to adapt in simple as well as in more complex speech interaction contexts. Moreover, the enjoyment related to music training and more specifically the fact of being actively involved in synchronizing with somebody during 30 min, may also have enhanced children's engagement in the interactional task and improved their consistency after rhythmic training (Kirschner and Tomasello, 2009).

5. Conclusion, limitations and perspectives

This study presented a new task which allows for examining temporal adaptation in children at verbal age in an interactive situation with a virtual partner. Results showed the relevance of rhythmic musical training to improve temporal skills in children with HL. The future development of this task, including the use of a more adaptive virtual agent and more complex utterances going beyond the use of single words, will help us to better understand on which level of conversational coordination rhythmic musical training can impact the most (Phillips-Silver and Keller, 2012). As our study was confined to a very periodic structure of turns, future investigations may extend to other time-sensitive phenomena in less constrained speech interaction, such as speech convergence (Pardo, 2006) which could benefit interlocutors by mutually reducing cognitive load and facilitating prosocial behavior (Manson et al., 2013).

As the musical training session was conceived to focus on rhythmic skills and entrainment, it may have enhanced not only predictive skills, but also auditory processing in general. This may, in turn, have enhanced encoding of acoustic cues such as pitch, intensity and duration, that characterize stress patterns (Torppa et al., 2014). Overall, this work supports the idea that musical stimulation in general and sensori-motor rhythmic training in particular can be integrated in speech and language rehabilitation of children with HL. Further studies should also address a direct comparison of children with NH and HL, which was beyond the scope of this work. Our results are also encouraging from a clinical perspective. Speech therapists with an appropriate focused musical training could develop simple but recurrent rhythmic activities with their patients with HL. Moreover, these rhythmic activities could also be carried out in a group, which was not feasible in the context of this protocol that required individual testing immediately after the training session.

Acknowledgements

This work was supported by the Brain and Language Research Institute (BLRI, ANR-11-LABEX-0036 to C.H., S.F. and D.S.), the European Union Seventh Framework Program (FP7-PEOPLE-2012-IEF, n° 327586 to S.F.), and by LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative (to S.F.).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.heares.2017.05.006>.

References

- Abney, D.H., Paxton, A., Dale, R., Kello, C.T., 2014. Complexity matching in dyadic conversation. *J. Exp. Psychol. Gen.* 143, 2304–2315. <http://dx.doi.org/10.1037/xge0000021>.
- Andrews, G., Howie, P.M., Dozsa, M., Guitar, B.E., 1982. Stuttering. *J. Speech Lang. Hear. Res.* 25, 208. <http://dx.doi.org/10.1044/jshr.2502.208>.
- Bebko, J.M., Calderon, R., Treder, R., 2003. The language proficiency Profile-2: assessment of the global communication skills of deaf children across languages and modalities of expression. *J. Deaf Stud. Deaf Educ.* 8, 438–451. <http://dx.doi.org/10.1093/deafed/eng034>.
- Beňuš, S., Gravano, A., Hirschberg, J., 2011. Pragmatic aspects of temporal accommodation in turn-taking. *J. Pragmat.* 43, 3001–3027. <http://dx.doi.org/10.1016/j.pragma.2011.05.011>.
- Berens, P., 2009. CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* 31, 1–21. <http://dx.doi.org/10.1002/wics.10>.
- Cannard, C., Blaye, A., Scheuner, N., Bonthoux, F., 2005. Picture naming in 3- to 8-year-old French children: methodological considerations for name agreement. *Behav. Res. Methods* 37, 417–425. <http://dx.doi.org/10.3758/BF03192710>.
- Casillas, M., Frank, M.C., 2013. The development of predictive processes in children's discourse understanding. In: *Proc. 35th Annu. Meet. Cogn. Sci. Soc.*, pp. 299–304.
- Cason, N., Astésano, C., Schön, D., 2015a. Bridging music and speech rhythm: rhythmic priming and audio-motor training affect speech perception. *Acta Psychol. (Amst)* 155, 43–50. <http://dx.doi.org/10.1016/j.actpsy.2014.12.002>.
- Cason, N., Hidalgo, C., Isoard, F., Roman, S., Schön, D., 2015b. Rhythmic priming enhances speech production abilities: evidence from prelingually deaf children. *Neuropsychology* 29, 102–107.
- Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. <http://dx.doi.org/10.1093/cercor/bhn042>.
- Cummins, F., Port, R., 1998. Rhythmic constraints on stress timing in English. *J. Phon.* 26, 145–171. <http://dx.doi.org/10.1006/jpho.1998.0070>.
- Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., Schön, D., 2015. Music training increases phonological awareness and reading skills in developmental dyslexia: a randomized control trial. *PLoS One* 10, e0138715. <http://dx.doi.org/10.1371/journal.pone.0138715>.
- Fujii, S., Wan, C.Y., 2014. The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Front. Hum. Neurosci.* 8, 1–15. <http://dx.doi.org/10.3389/fnhum.2014.00777>.
- Garrod, S., Anderson, A., 1987. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition* 27, 181–218. [http://dx.doi.org/10.1016/0010-0277\(87\)90018-7](http://dx.doi.org/10.1016/0010-0277(87)90018-7).
- Garrod, S., Pickering, M.J., 2015. The use of content and timing to predict turn transitions. *Front. Psychol.* 6, 1–12. <http://dx.doi.org/10.3389/fpsyg.2015.00751>.
- Garrod, S., Pickering, M.J., 2004. Why is conversation so easy? *Trends Cogn. Sci.* 8, 8–11. <http://dx.doi.org/10.1016/j.tics.2003.10.016>.
- Geiser, E., Sandmann, P., Jäncke, L., Meyer, M., 2010. Refinement of metre perception - training increases hierarchical metre processing. *Eur. J. Neurosci.* 32, 1979–1985. <http://dx.doi.org/10.1111/j.1460-9568.2010.07462.x>.
- Giles, H., Coupland, N., Coupland, J., 1991. Accommodation theory: communication, context, and consequence. In: Giles, H., Coupland, J., Coupland, N. (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press, pp. 1–68.
- Hannon, E.E., Trehub, S.E., 2005. Metrical categories in infancy and adulthood. *Psychol. Sci.* 16, 48–55. <http://dx.doi.org/10.1111/j.0956-7976.2005.00779.x>.
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6, 1–12. <http://dx.doi.org/10.3389/fpsyg.2015.01492>.
- Himberg, T., Hirvankari, L., Mandel, A., Hari, R., 2015. Word-by-word entrainment of speech rhythm during joint story building. *Front. Psychol.* 6, 1–6. <http://dx.doi.org/10.3389/fpsyg.2015.00797>.
- Innes-Brown, H., Marozeau, J.P., Storey, C.M., Blamey, P.J., 2013. Tone, rhythm, and timbre perception in school-age children using cochlear implants and hearing aids. *J. Am. Acad. Audiol.* 24, 789–806. <http://dx.doi.org/10.3766/jaaa.24.9.4>.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C.L., Jasnow, M.D., 2001. Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child. Dev.* 66, 1–132. <http://dx.doi.org/10.2307/3181589>.
- Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., Yamaguchi, Y., 2013. Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3, 1–8. <http://dx.doi.org/10.1038/srep01692>.
- Kirschner, S., Tomasello, M., 2009. Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child. Psychol.* 102, 299–314. <http://dx.doi.org/10.1016/j.jecp.2008.07.005>.
- Liss, J.M., White, L., Mattys, S.L., Lansford, K., Lotto, A.J., Spitzer, S.M., Caviness, J.N., 2009. Quantifying speech rhythm abnormalities in the dysarthrias. *J. Speech. Lang. Hear. Res.* 52, 1334–1352. [http://dx.doi.org/10.1044/1092-4388\(2009\)08-0208](http://dx.doi.org/10.1044/1092-4388(2009)08-0208).
- Manson, J.H., Bryant, G.A., Gervais, M.M., Kline, M.A., 2013. Convergence of speech rate in conversation predicts cooperation. *Evol. Hum. Behav.* 34, 419–426. <http://dx.doi.org/10.1016/j.evolhumbehav.2013.08.001>.
- Marie, C., Magne, C., Besson, M., 2011. Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305. <http://dx.doi.org/10.1162/jocn.2010.21413>.
- Most, T., Shina-August, E., Meilijson, S., 2010. Pragmatic abilities of children with hearing loss using cochlear implants or hearing AIDS compared to hearing children. *J. Deaf Stud. Deaf Educ.* 15, 422–437. <http://dx.doi.org/10.1093/deafed/eng032>.
- Natale, M., 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *J. Personal. Soc.* 1, 790–804. <http://dx.doi.org/10.1037/0022-3514.32.5.790>.
- Pardo, J.S., 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382. <http://dx.doi.org/10.1121/1.2178720>.
- Petersen, B., Weed, E., Sandmann, P., Brattico, E., Hansen, M., Sørensen, S.D., Vuust, P., 2015. Brain responses to musical feature changes in adolescent cochlear implant users. *Front. Hum. Neurosci.* 9, 7. <http://dx.doi.org/10.3389/fnhum.2015.00007>.
- Phillips-Silver, J., Keller, P.E., 2012. Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6, 1–11. <http://dx.doi.org/10.3389/fnhum.2012.00026>.
- Phillips-Silver, J., Toivainen, P., Gosselin, N., Turgeon, C., Lepore, F., Peretz, I., 2015. Cochlear implant users move in time to the beat of drum music. *Hear. Res.* 321, 25–34. <http://dx.doi.org/10.1016/j.heares.2014.12.007>.
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Lécuyer, L., Kotz, S.A., Tillmann, B., 2013. Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* 27, 121–131. <http://dx.doi.org/10.1037/a0031277>.
- Repp, B.H., 2005. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. <http://dx.doi.org/10.3758/BF03206433>.
- Roy, A.T., Scattergood-Keppner, L., Carver, C., Jiradejvong, P., Butler, C., Limb, C.J., 2014. Evaluation of a test battery to assess perception of music in children with cochlear implants. *JAMA Otolaryngol. Head. Neck Surg.* 140, 1–8. <http://dx.doi.org/10.1001/jamaoto.2014.341>.
- Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. <http://dx.doi.org/10.1016/j.conb.2010.02.010>.
- Schubotz, R.I., 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218. <http://dx.doi.org/10.1016/j.tics.2007.02.006>.
- Späth, M., Aichert, I., Ceballos-Baumann, A.O., Wagner-Sonntag, E., Miller, N., Ziegler, W., 2016. Entraining with another person's speech rhythm: evidence from healthy speakers and individuals with Parkinson's disease. *Clin. Linguist. Phon.* 30, 1–18. <http://dx.doi.org/10.3109/02699206.2015.1115129>.
- Stabej, K.K., Smid, L., Gros, A., Zargi, M., Kosir, A., Vatovec, J., 2012. The music perception abilities of prelingually deaf children with cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* 76, 1392–1400. <http://dx.doi.org/10.1016/j.ijporl.2012.07.004>.
- Stahl, B., Kotz, S.A., Henseler, I., Turner, R., Geyer, S., 2011. Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain* 134, 3083–3093. <http://dx.doi.org/10.1093/brain/awr240>.
- Street, R., 1984. Speech convergence and speech evaluation in fact-finding interviews. *Hum. Commun. Res.* 11, 139–169. <http://dx.doi.org/10.1111/j.1468-2958.1984.tb00043.x>.
- Thompson, W.F., Schellenberg, E.G., Husain, G., 2001. Arousal, mood, and the Mozart effect. *Psychol. Sci.* 12, 248–251. <http://dx.doi.org/10.1111/1467-9280.00345>.
- Tierney, A., Kraus, N., 2013. Music Training for the Development of Reading Skills. Progress in brain research, first ed. Elsevier B.V. <http://dx.doi.org/10.1016/B978-0-444-63327-9.00008-4>.
- Toe, D.M., Paatsch, L.E., 2013. The conversational skills of school-aged children with cochlear implants. *Cochlear Implants Int.* 14, 67–79. <http://dx.doi.org/10.1179/1754762812Y.0000000002>.
- Torppa, R., Faulkner, A., Huotilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., Vainio, M., 2014. The perception of prosody and associated auditory cues in early-implanted children: the role of auditory working memory and musical activities. *Int. J. Audiol.* 53, 182–191. <http://dx.doi.org/10.3109/14992027.2013.872302>.
- Torppa, R., Salo, E., Makkonen, T., Loimo, H., Pykäläinen, J., Lipsanen, J., Faulkner, A., Huotilainen, M., 2012. Cortical processing of musical sounds in children with cochlear implants. *Clin. Neurophysiol.* 123, 1966–1979. <http://dx.doi.org/10.1016/j.clinph.2012.03.008>.
- Tye-Murray, N., 2003. Conversational fluency of children who use cochlear implants. *Ear Hear* 24, 825–895. <http://dx.doi.org/10.1097/01.AUD.0000051691.33869.EC>.
- van Wieringen, A., Wouters, J., 2015. What can we expect of normally-developing children implanted at a young age with respect to their auditory, linguistic and cognitive skills? *Hear. Res.* 322, 171–179. <http://dx.doi.org/10.1016/j.heares.2014.09.002>.
- Volman, M.J., Geuze, R.H., 2000. Temporal stability of rhythmic tapping “on” and “off the beat”: a developmental study. *Psychol. Res.* 63, 62–69. <http://dx.doi.org/10.1007/PL00008168>.
- Vuust, P., Wittek, M.A.G., 2014. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5,

- 1–14. <http://dx.doi.org/10.3389/fpsyg.2014.01111>.
- Wan, C.Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., Norton, A., Schlaug, G., 2011. Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: a proof of concept study. *PLoS One* 6, 1–7. <http://dx.doi.org/10.1371/journal.pone.0025505>.
- Warner, R.M., Malloy, D., Schneider, K., Knoth, R., Wilder, B., 1987. Rhythmic organization of social interaction and observer ratings of positive affect and involvement. *J. Nonverbal Behav.* 11, 57–74. <http://dx.doi.org/10.1007/BF00990958>.

Expérience n°3



Rhythmic training improves temporal conversational abilities in children with hearing loss

Journal:	<i>Journal of Speech, Language, and Hearing Research</i>
Manuscript ID	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hidalgo, Céline; Laboratoire Parole et Langage, Sciences du Langage; Institut de Neurosciences des Systèmes, Neurosciences Pesnot--Lerousseau, Jacques; Institut de Neurosciences des Systèmes, Neurosciences Marquis, Patrick; Institut de Neurosciences des Systèmes, Neurosciences Roman, Stéphane; Hopital de la Timone, ORL pédiatrique et Chirurgie Cervico-Faciale; Institut de Neurosciences des Systèmes, Neurosciences Nguyen, Noël; Laboratoire Parole et Langage, Sciences du Langage Schon, Daniele; Institut de Neurosciences des Systèmes, Neurosciences
Keywords:	Cochlear Implants, Intervention, Speech, Neuroimaging, Communication

SCHOLARONE™
Manuscripts

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4

5 **Rhythmic training improves temporal conversational abilities**
6 **in children with hearing loss**
7

8 Céline Hidalgo^{1,2}, Jacques Pesnot-Lerousseau², Patrick Marquis², Stéphane Roman^{2,3}, Noël
9 Nguyen¹, Daniele Schön²
10
11
12

13 ¹Aix Marseille Univ, CNRS, LPL, Aix-en Provence, France

14 ²Aix Marseille Univ, Inserm, INS, Inst Neurosci Syst, Marseille, France

15 ³Pediatric Otolaryngology department, La Timone Children's Hospital (APHM), Marseille, France
16
17
18
19
20
21
22

23 Corresponding author: Céline Hidalgo, celine.hidalgo@univ-amu.fr, Laboratoire Parole et
24 Langage, UMR 7309, CNRS, Aix-Marseille University, 5 Avenue Pasteur, 13100 Aix-en-
25 Provence, France
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3

4 **Abstract**
5
6

7 During conversation, the ability to adapt to the partner's speech is crucial for an optimal
8 communicative exchange. In this study, we investigate temporal adaptation capacities of
9 children with normal hearing and children with cochlear implants (CIs) and/or hearing aids
10 (HAs), during verbal exchange. We also address the question of the efficiency of a rhythmic
11 training on temporal adaptation during speech interaction in children with hearing loss. We
12 recorded EEG in children while they named pictures delivered on a screen, in alternation with
13 a virtual partner. We manipulated the virtual partner's speech rate (*fast* versus *slow*) and the
14 regularity of alternation (*regular* versus *irregular*). The group of children with normal hearing
15 was tested once and the group of children with hearing loss twice: once after 30 minutes of
16 auditory training and once after 30 minutes of rhythmic training. Both groups of children
17 adjusted their speech rate to that of the virtual partner and were sensitive to the regularity of
18 alternation with a less accurate performance following irregular turns. Moreover, irregular
19 turns elicited an MMN-like response in both groups, showing a detection of temporal
20 deviancy. Notably, the MMN-like amplitude positively correlated with accuracy in the
21 alternation task. In children with hearing loss, the effect was more pronounced and long-
22 lasting following rhythmic training compared with auditory training. These results are
23 discussed in terms of adaptation abilities in speech interaction and suggest the use of rhythmic
24 training to improve conversational temporal skills of children with hearing loss.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 **Keywords**
47

48 Cochlear implant, rehabilitation, speech, EEG, temporal predictions
49
50
51
52
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3

4 Introduction

5
6 The most natural and common usage of speech is in conversation. Interactions between
7
8 interlocutors heavily rely on anticipatory processes in order to optimize communicative
9
10 exchanges and minimize interruptions (Pickering & Garrod, 2013; Levinson & Torreira,
11
12 2015). In fact, as it is the case in any joint action implying two or more individuals, a major
13
14 key of the success of an optimal communicative exchange is the ability for each interlocutor
15
16 to temporally predict her/his partner's actions (Sebanz & Knoblich, 2009; Kourtis, Sebanz, &
17
18 Knoblich, 2013). For instance, anticipating the end of the speech turn on the basis of the
19
20 speech rate allows turn-taking at an optimal pace (Wilson & Wilson, 2005). An anticipation
21
22 of turn-ends allows avoiding speech overlap of the interlocutors as well as over-long silences
23
24 that may break the fluency of the exchange. Moreover, it may also allow a better perception
25
26 of the grouping and prominence of speech that may in turn facilitate speech spectral
27
28 processing (Cummins & Port, 1998; Port, 2003; Peelle & Davis, 2012).
29
30

31
32 This ability to readily take turns is possibly facilitated by the fact that interlocutors tend to
33
34 adopt a common rhythm, based on shared syllable timing (Street, 1984; Himberg, Hirvenkari,
35
36 Mandel, & Hari, 2015). This phenomenon is considered as taking place automatically and is
37
38 often referred to as between-speaker rhythmic convergence.
39

40
41 Recent studies have focused on how these anticipatory skills develop during infancy. It has
42
43 been shown that 5-month-old babies already engage in turn-taking with their mothers, with
44
45 few overlaps and short turn-taking windows (Hilbrink, Gattis, & Levinson, 2015).
46

47
48 Children with a prelingual hearing loss (HL), whether equipped with cochlear implant or
49
50 conventional prostheses, are particularly vulnerable in terms of temporal prediction abilities
51
52 because of the early auditory deprivation. This has a direct impact upon speech perception
53
54 abilities, especially in acoustically-degraded environmental conditions, as well as on the
55
56 mastery of certain communication skills. More precisely, children with HL have difficulties in
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 imitating the speech rate of their interlocutor (i.e. temporal convergence) and this goes along
4 with reduced speech intelligibility (Freeman & Pisoni, 2017). These children are also less
5 sensitive than normal-hearing children to the temporal regularity of stress distribution in
6 speech in an interactive verbal task (Hidalgo, Falk, & Schön, 2017). Adolescents with CI both
7 use different strategies for repairing communication failures and are less sensitive to prosodic
8 cues relevant to spoken discourse understanding (Holt, Demuth, & Yuen, 2016; Holt, Yuen,
9 & Demuth, 2017) compared to normal-hearing peers.

10
11 In a rehabilitation perspective, it seems essential to take into account the alteration of this type
12 of communication skills in hearing-impaired children in order to improve the quality of their
13 verbal interactions as well as the development of their social relations (Hoffman, Quittner, &
14 Cejas, 2014; Manson, Bryant, Gervais, & Kline, 2013; Wong et al., 2017). An intervention
15 aimed at improving temporal prediction abilities based on rhythm training seems particularly
16 relevant.

17
18 Indeed, in a sentence repetition task, presenting a musical rhythm before each sentence
19 improves hearing-impaired children's accuracy, even more so when the rhythm of the musical
20 prime matches that of the sentence (Cason, Hidalgo, Isoard, Roman, & Schön,
21 2015). Moreover, we (Hidalgo et al., 2017) recently showed that following 30 minutes of
22 rhythmic training (compared to a control condition), children with HL show an enhanced
23 capacity to accommodate to a temporally predictable flow of turn exchanges in an
24 interactional setting. These results fit well with the literature showing the beneficial effect of
25 musical expertise, particularly in the rhythmical domain, on sensitivity to speech rhythm
26 (Marie, Magne, & Besson, 2011; Moritz, Yampolsky, Papadelis, Thomson, & Wolf, 2013;
27 Magne, Jordan, & Gordon, 2016), and that of rhythmic stimulation on speech processing
28 (Cason, Astésano, & Schön, 2015; Cason & Schön, 2012; Chern, Tillmann, Vaughan,
29 Gordon, & Gordon, 2017), both in normal people and people with a variety of language

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 disorders (Kotz, Gunter, & Wonneberger, 2005; Przybylski et al., 2013; Cason, Hidalgo, et
4 al., 2015; Flaugnacco et al., 2015; Bedoin, Brisseau, Molinier, Roch, & Tillmann, 2016;
5
6 Schön & Tillmann, 2015).

7
8
9
10 In the light of the above-mentioned studies illustrating the beneficial effect of rhythmic
11 stimulation on temporal predictions and the processing of auditory information in both adults
12 and children (for a review, see Haegens & Zion Golumbic, 2017), we conducted a combined
13 EEG and behavioral study to measure children's temporal adaptation capacities and to
14 examine the influence of musical rhythmic training on speech perception and production
15 abilities in children with HL.
16
17

18
19 More precisely, we ran two experiments. In Experiment 1, children with normal hearing (NH)
20 were asked to perform a picture-naming task in alternation with a virtual partner (adapted
21 from Hidalgo et al., 2017). The speech rate of the virtual partner could be slow or fast and the
22 temporal structure of turns could be regular or irregular. We performed acoustic analyses of
23 speech production to determine to what extent children converged towards the virtual
24 partner's (VP) speech rate as well as to measure accuracy and consistency of word stress
25 placement across regular and irregular turns. Event-related potentials were used to assess the
26 sensitivity to turn regularity.
27
28
29
30
31
32
33
34
35
36
37
38
39

40 In Experiment 2, children with HL, wearing either cochlear implants (CIs) or hearing aids
41 (HAs), did the picture-naming task twice: once following 30 minutes of rhythmic musical
42 training, and once following 30 minutes of auditory training. We hypothesized that the
43 rhythmic training would improve the ability of children with HL to predict conversational
44 dynamics which in turn would facilitate speech perception and production.
45
46
47
48
49

50
51 Overall, the aim of this study was twofold. Firstly, we used a turn-taking task to assess in
52 normal hearing (NH) children the *speech rate convergence* ability as well as the
53 *accommodation* ability. to adapt to temporal deviancy in the regularity of the turns.
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 Convergence was assessed by estimating changes of the speech rate in the fast compared to
5
6 the slow condition. Accommodation was assessed by estimating changes in alternation in the
7
8 irregular compared to the regular condition. Secondly, we tested the hypothesis that a single
9
10 session of rhythmic training, by improving temporal prediction, may facilitate perception,
11
12 convergence and accommodation to the temporal deviancy in a conversational context in
13
14 children with HL.
15
16
17
18

19 Experiment 1

20 **Material and Method**

21 **Participants**

22
23
24
25 Twenty-two children with normal hearing (NH) (13 girls) were recruited from an *elementary*
26
27 *school* in Marseille (France). They were aged from 6 to 10 years (*mean* = 102.6 months, *SD* =
28
29 18.5 months) and were all French native speakers without any known visual, speech,
30
31 cognitive or hearing disorder. This experiment was approved by the Ethics Committee Sud
32
33 Méditerranée I (n°ID RCB: 2015-A01490-49). All parents signed a consent form.
34
35

36 **Stimuli**

37
38 For the alternation naming task, we selected 46 pictures from the French BD2I database
39
40 (Cannard, Blaye, Scheuner, & Bonthoux, 2005). These pictures were named correctly by over
41
42 90% of a normative sample of 3 to 8-year-old children (mean score = 98.8%, *SD*=1.6%).
43

44
45 46 picture names (disyllabic words) were recorded in a soundproof booth by a male native
46
47 French speaker (henceforth, the virtual partner, VP). Four words were also recorded by an 8-
48
49 year-old male native French speaker to be used as an example of the alternating procedure at
50
51 the beginning of each block.
52

53
54 We performed an automatic word segmentation of the VP's recording using SPPAS (Bigi,
55
56 2015, 1.8.0 version) and PRAAT (Boersma & Weenink, 2012). All words were then
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 manipulated to obtain a fast (80% of initial word duration) and a slow (120%) version for
4
5 each of them, using a wavelet-based method for audio dilation and compression (Pallone et al,
6
7 1999). For both the fast and slow versions, p-centers, taken as rhythmic perceptual anchor
8
9 points (Morton, Marcus, & Frankish, 1976), were semi-automatically located on the vowel of
10
11 the second, stressed syllable, as the 2/3rd of the vowel energy rise (Cummins & Port, 1998).
12
13

14 Procedure

15
16 Before the beginning of the picture-naming task, children were familiarized with the pictures
17
18 and the experimental procedure to make sure that the words produced for each picture both
19
20 were the expected ones and were pronounced in an intelligible manner. Then, children were
21
22 equipped with a 21-channel electrode cap (International 10/20 system sites) and were seated
23
24 in a Faraday booth in front of a computer screen. They were asked to name the pictures
25
26 displayed on the screen, in alternation with a virtual partner (Figure 1).
27
28

29 =====

30
31
32 PLEASE INSERT FIGURE 1 AROUND HERE

33
34 =====

35
36
37 The experiment consisted of two separate blocks, one with a slow and one with a fast VP
38
39 speech rate. Each block started with 4 pairs of pictures that were named by the adult virtual
40
41 partner in perfect alternation with the voice of a pre-recorded 8-year-old boy in order to
42
43 establish, implicitly, a regular pace of alternation. Next, 132 pictures were presented
44
45 successively on the screen and the child was instructed to name the pictures in alternation
46
47 with the VP. Care was taken to avoid semantic or phonological priming between successive
48
49 words. A given word pair (VP-child) never appeared twice. The child was praised and
50
51 encouraged before and between the two blocks. A short pause was given between blocks. The
52
53 duration of each block was around 5 minutes.
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3 While, as stated above, most utterances of the VP were presented at a regular temporal
4 interval of 2000ms, ~20% of the words of the VP were anticipated at 1600ms (in both blocks)
5
6 and pseudo-randomly distributed across trials.
7
8

9
10 This resulted in a 2 by 2 factorial design manipulating the Regularity of alternation (*regular*
11 or *irregular*) and Speech Rate (*slow* or *fast*) across blocks. The order of block presentation
12 (*fast* versus *slow*) was counterbalanced across participants.
13
14

15
16 Stimuli delivery was programmed using the software Presentation (Neurobehavioral Systems
17 18.0). Pictures were displayed on a computer (DELL Precision T1700, screen resolution 1366
18 x 768). The voice of the VP was delivered through a 2040 YAMAHA amplifier and two NS
19 1020 Studio YAMAHA loudspeakers. The output sound volume was adjusted at a
20 comfortable level for each child. The child's speech was recorded with a lapel microphone
21 (Sennheiser) linked to a ZOOM H4n numeric recorder at a 44 kHz sampling rate.
22
23

24
25 For every child, we also assessed rhythmic and neuropsychological abilities. Rhythmic
26 assessment included tapping along a 90 pulses/minute metronome (henceforth paced tapping)
27 test, tapping in syncopation along a 55 pulses/minute metronome (henceforth syncopated
28 tapping) and reproduction of 10 rhythms (3 to 8 sounds each) (henceforth rhythmic patterns
29 reproduction). Children tapped onto a cardboard box containing a ZOOM H4n digital
30 recorder. Neuropsychological tests were issued from NEPSY and WISC IV. These tests
31 included digits direct and indirect span test and an auditory attention test requiring to place a
32 token in a box each time the child heard "red" among other words.
33
34
35
36
37
38
39
40
41
42
43
44
45

46 **Behavioral data processing and statistical analysis**

47
48 Children's convergence abilities were evaluated by comparing word duration (in ms) across
49 the two speech rates (fast and slow).
50
51

52
53 Children's accommodation capacities were determined by the consistency of the children's p-
54 center placement across trials and the accuracy (temporal asynchrony) relative to the VP's
55
56
57

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 stress placement. First, we extracted the temporal interval (in ms) between the p-centers of the
5
6 VP and the p-centers of children (Inter-Word-Interval, IWI). Then we computed circular
7
8 statistics (CircStats, Berens, 2009) on the IWI in order to express the children's p-center
9
10 placement as a vector in a circular space relative to the expected moment in time (half of the
11
12 inter-trial interval of the virtual partner) (statistics also used in Hidalgo et al., 2017 &
13
14 Himberg et al., 2015). P-center placement consistency was measured by the vector length, and
15
16 accuracy by the vector angle. Words produced with an error, a hesitation or simultaneously
17
18 with the virtual partner were excluded from all analyses, resulting in rejection of
19
20 approximately 30% of the children's productions.
21
22

23 We used R (R Core Team, 2013) and the lme4 package (Bates, Maechler & Bolker,
24
25 2013) to fit linear mixed models on single trial data for both word duration and vector angle.
26
27 All models were built with participants as random effect. Statistical significance of the fixed
28
29 effects was assessed by model comparison on AIC, thus arbitrating between complexity and
30
31 explanatory power of the models. Normality and homoscedasticity of the residuals were
32
33 systematically visually inspected. Reported p-values are Satterthwaite approximations
34
35 obtained with the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2015). As
36
37 consistency is always positive and not normally distributed, we applied a logit transform
38
39 before model fitting (Falk, Müller & Dalla Bella, 2015). Furthermore, random effects were
40
41 dropped for consistency, as this summary statistics (estimated by vector length) gives only
42
43 one value per participant and condition.
44
45

46 In the *paced* and *syncopated tapping* tasks, tapping times were extracted using Adobe
47
48 Audition (2.0) and circular analyses were used to compute the children's consistency and
49
50 accuracy.
51
52

53 In the *rhythmic patterns reproduction*, the anonymized children's performances were scored
54
55 from 1 to 9 (1: no reproduction, 9: perfect performance) by one judge (the last author) with
56
57

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 expertise in music using a blind procedure. The average score across trials was used for
4
5 further analyses.
6

7 Correlations were computed using Spearman's rank, reported with 95% confidence intervals,
8
9 between rhythmic measures (paced tapping, and rhythmic patterns reproduction), behavioural
10
11 and electrophysiological responses during the turn-taking task.
12
13

14 **Electrophysiological data processing and statistical analysis**

15 EEG signal processing was done using EEGLAB (Delorme & Makeig, 2004) and custom
16
17 Matlab scripts. Continuous data were filtered using a high-pass filter (cutoff frequency 1 Hz)
18
19 and major artifacts rejected. Independent Component Analysis (ICA) was used to remove
20
21 physiological artifacts such as eye blinks and muscular activity. For the Event-Related
22
23 Potentials (ERPs) analysis, data were segmented into 900ms epochs starting 100ms before the
24
25 presentation of the VP's words onsets. Further rejection of remaining artifacts was done on
26
27 the basis of visual inspection of each epoch and was always lower than 10% of the total
28
29 number of epochs in a given condition.
30
31
32
33
34
35

36 We conducted two types of statistical analysis. The first analysis was run on the peak of the
37
38 MMN using a classical pairwise comparison. First, a temporal window for peak search was
39
40 defined using the average of all trials. Then, for *regular* and *irregular* conditions, the peak
41
42 amplitude was computed as the mean value ± 40 ms around the individual peak. A Region of
43
44 Interest (ROI) was defined on the basis of the topography of the average of all conditions,
45
46 using a hierarchical clustering technique. The MMN-like ROI comprised 'P3', 'P4', 'O1',
47
48 'O2', 'P7', 'P8', and 'Pz' (see Figure 4A). The second analysis consisted of a cluster-based
49
50 analysis in the time domain on the same ROI to analyze the differences between *regular* and
51
52 *irregular* conditions. Significance was assessed using non-parametric pairwise two-tailed
53
54 permutation tests that provide corrected p-values for multiple comparisons. For each
55
56
57
58
59
60

Rhythmic training improves temporal conversational abilities in children with hearing loss permutation ($N = 8,191$), time clusters were defined on the basis of temporal adjacency by regrouping samples whose T-statistic was larger than 1.73, *i.e.* $p\text{-value} \leq 0.05$ for a t-test with 19 degrees of freedom. Cluster-level statistics were then calculated by taking the sum of the T-values within the cluster. Only temporal clusters with corrected $p\text{-values} \leq 0.05$ are reported.

Results

Behavioral results. The speech rate of the VP affected children's word duration ($\beta = -33.37$, $SE = 3.76$, $t = -8.85$, $p < .001$). More precisely, children's words lasted longer in the *Slow* compared to the *Fast* condition (see Figure 2A).

The analyses of stress placement focused on *consistency*, *i.e.* whether the child placed the stress consistently in the same time window across trials, and *accuracy*, *i.e.* how close the stress was placed with respect to the expected time point.

The regularity of alternation influenced children's consistency and accuracy in stress placement. Children placed word p-centers more consistently in the *Regular* compared to the *Irregular* condition ($\beta = 0.74$, $SE = 0.28$, $t = 2.667$, $p < .01$) (see Figure 2B). They also placed word stresses close to the expected time point in the *Regular* trials while they were late in *Irregular* trials ($\beta = -0.81$, $SE = 0.03$, $t = -24.09$, $p < .001$, Figure 2B).

=====

PLEASE INSERT FIGURE 2 AROUND HERE

=====

Electrophysiological results. Results showed a MMN-like effect following Irregular trials ($ES = -3.7\mu V$, $t(1,19) = -5.5834$, $p < 0.001$). This effect was significant across a long-lasting time cluster, from 124 to 492ms after stimulus onset ($p < 0.01$). Correlation analysis on the peak amplitude of the MMN-like effect and the vector length showed that children who were more

Rhythmic training improves temporal conversational abilities in children with hearing loss sensitive to the temporal deviancy of the virtual partner, *i.e.* with a large MMN-like effect, were also more consistent in the turn-taking task, *i.e.* with a high vector length (*Regular* $r = .58$, CI [.85, .14]; *Irregular* $r = .44$, CI [.73,.003]). MMN-like amplitude (MMN-like effect size in the significant cluster) was also positively correlated with the accuracy in stress placement during the *Irregular* intervals ($r = .53$; CI [.09,.85]). Finally, we also found a positive correlation between children consistency in the tapping task and the turn-taking task during *Regular* intervals ($r = .50$ CI [.07, .75]). The reproduction of rhythmic patterns did not show clear correlations with any of the other variables.

Experiment 2

Material and methods

Participants

Sixteen children with HL (7 girls) wearing bilateral or unilateral cochlear implants (CIs) and/or Hearing Aids (HAs) (see Table 1 for details) were recruited, mostly via the ENT paediatric department at the Timone Hospital in Marseilles. All but two were enrolled in a mainstream primary school. They were aged from 6 to 10 years (*mean* = 102 months, *SD* = 15.6 months) and were all native speakers of French without any reported visual or associated speech and cognitive impairment. Data of two participants were excluded because of poor EEG data quality. This experiment was approved by the Ethics Committee Sud Méditerranée I (n°ID RCB: 2015-A01490-49).

=====

PLEASE INSERT TABLE 1 AROUND HERE

=====

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4
5

6 **Stimuli and procedure**

7
8 The stimuli, task and procedure were identical to Experiment 1 with the exception that
9
10 children with HL performed the task twice: once after 30 minutes of a speech-therapy like
11
12 auditory training, and once after 30 minutes of a session of active rhythm. The two sessions
13
14 were always separated by at least one week and their order was counterbalanced across
15
16 children.
17

18
19 The rhythmic training comprised the following exercises: Follow the beat (balance the arms
20
21 and walk with a metronome). Structure the beat into meter (listen and tap the strong beat with
22
23 the feet and feel and then tap other beats clapping hands with therapist). Learn a new rhythm
24
25 (listen to music and find a repetitive rhythm, move to it and tap using claves). Follow metric
26
27 changes (listen to music and change body movement type with metrical changes). Body
28
29 tapping (in absence of an external stimulus clap hands, tap chest, stop and keep internal beat).
30
31 Beat boxing (produce rhythmic patterns with mouth, alone and with therapist).
32
33

34
35 The auditory training comprised the following exercises: Timbre recognition across categories
36
37 (recognize sounds belonging to different categories: animals, musical instruments,
38
39 environmental sounds). Timbre recognition within categories (recognize sounds belonging to
40
41 the same category: animals, musical instruments, environmental sounds). Sound sequences
42
43 recognition across categories (recognize sequences of increasing length comprising sounds
44
45 belonging to different categories). Sound sequences recognition within categories (recognize
46
47 sequences of increasing length comprising sounds belonging to a same category). The level of
48
49 difficulty in these exercises was manipulated by reducing the timbral difference between
50
51 sounds.
52
53

54 **Processing and statistical analyses**

55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3 Behavioral analyses were identical to Experiment 1. We tested the effects of speech rate,
4
5 regularity and training as well as all the interactions.
6
7

8 EEG data processing was also identical to Experiment 1 with the only exception that a second
9
10 ICA was run on EEG epoched data to isolate and remove the potential artefacts generated by
11
12 the cochlear implant or hearing aids (Campos Viola et al., 2012). Spatial ROI for MMN-like
13
14 effect was the same as the NH children. Statistical analyses were also the same, with the only
15
16 exception that the threshold for cluster statistic in time was set to 1.77, *i.e.* $p\text{-value} \leq 0.05$ for
17
18 a t-test with 13 degrees of freedom.
19
20

21 In order to reduce the number of comparisons we only computed correlations on the
22
23 correlated variables found in the NH group analysis, *i.e.* those with $r > .40$. Data from the
24
25 Rhythmic and the Auditory sessions were averaged.
26
27

28 **Results**

29
30
31 *Behavioral results.* As for NH children, the speech rate of the virtual partner affected the
32
33 speech rate of children with HL, in both conditions of stimulation (*auditory Training* and
34
35 *rhythmic Training*). More precisely, children's word durations were shorter in the *fast*
36
37 condition compared to the *slow* condition after both *Auditory training* ($\beta = -39.75$, $SE = 3.86$,
38
39 $t = -10.27$, $p < .001$) and *Rhythmic training* ($\beta = -42.24$, $SE = 3.76$, $t = -11.22$, $p < .001$). The
40
41 interaction term did not improve the fit of the model ($\beta = 2.22$, $SE = 5.67$, $t = 3.92$, $p = .695$)
42
43 (see Figure 3A and 3C).
44
45

46 =====

47
48
49 PLEASE INSERT FIGURE 3 AROUND HERE

50
51 =====

52
53
54 As for NH children, the regularity of alternation influenced stress placement consistency and
55
56 accuracy of children with HL. Children with HL placed the stress more consistently in the
57
58

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
 2
 3
 4 *regular* compared to the *irregular* trials ($F(2.53) = 5.75, p = .005$) (Figure 3B and 3D).
 5
 6 They also placed word stresses closer to the expected time point in the *regular* trials compared
 7
 8 to *irregular* trials ($\beta = -0.66, SE = 0.03, t = -10.27, p < .001$) (Figure 3B and 3D). While the
 9
 10 interaction between speech rate and regularity did not improve the model significantly ($p =$
 11
 12 0.5), the interaction between training type and regularity of alternation significantly improved
 13
 14 the fit of the model for accuracy data: the perturbation induced by irregular trials on accuracy
 15
 16 of stress placement was reduced following *Rhythmic training* ($\beta = -0.66, SE = 0.03, t = -17.61,$
 17
 18 $p < .001$) compared to *Auditory training* ($\beta = -0.79, SE = 0.03, t = -21.39, p < .001$ see Figure
 19
 20 3B).
 21
 22

23 *Electrophysiological results.* Analyses on the peak showed a MMN-like effect following both
 24
 25 *Rhythmic Training* and *Auditory Training* ($t(1, 47) = -6.4654, p < .0001$). The MMN peak
 26
 27 amplitude effect did not differ across the two conditions of stimulation ($t(1, 13) = 1.1778, p =$
 28
 29 0.26).
 30
 31

32 By contrast, clusters analysis revealed a long lasting MMN-like effect from 239 to 469ms
 33
 34 following *Rhythmic Training* ($p < .05$) (see figure 4B) while no cluster reached significance
 35
 36 following *Auditory Training* ($p > 0.4$) (see Figure 4C).
 37
 38

39 Correlation analyses revealed that children who were more sensitive to the temporal deviancy
 40
 41 of the virtual partner (MMN-like effect size in the significant cluster) were also more accurate
 42
 43 in the stress placement in the *irregular* turns ($r = .42$ CI $[-.03, .65]$).
 44
 45

46 =====

47
 48 PLEASE INSERT FIGURE 4 AROUND HERE
 49

50 =====

51 52 53 Discussion 54 55 56 57 58 59 60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3
4 The aim of this study was twofold. Firstly, we measured the speech rate convergence and
5
6 temporal accommodation abilities in 6 to 10-year-old children during a turn-taking task with a
7
8 virtual partner, and to bridge these abilities to the sensitivity to temporal deviancy as
9
10 measured by an MMN-like effect. Secondly, we assessed whether a single 30-minute session
11
12 of rhythmic training can enhance temporal prediction in children with hearing loss and, in
13
14 turn, affect speech temporal skills that are relevant for a fluent conversation.
15
16

17
18 Results show that both children with NH and children with HL converge towards the speech
19
20 rate of a virtual partner. Moreover, in both children groups, stress placement is more
21
22 consistent and accurate following regular VP turns compared to irregular turns. Most
23
24 importantly, the rhythmic training session improves the accuracy of children with HL in
25
26 placing the stress at irregular turns. In other words, rhythmic training seems to endow children
27
28 with greater flexibility in adapting to a sudden change in the temporal context. Finally, both
29
30 groups (NH-HL) show an MMN-like response to irregular turns of the virtual partner. This
31
32 response is larger in children with HL following rhythmic training compared to the same
33
34 children following auditory training. We will first discuss children convergence abilities by
35
36 analysing the effects of the virtual partner's speech rate on children speech, then the effect of
37
38 alternation regularity on children's temporal accommodation abilities and, in a second step,
39
40 we will discuss the effects of rhythmic training on temporal accommodation capacities of
41
42 children with HL.
43
44
45

46 **Speech rate convergence with a virtual partner**

47
48
49 One innovative aspect of this experimental design is that it allows, by manipulating the virtual
50
51 partner's speech rate, to estimate the speech rate convergence, namely the natural tendency to
52
53 minimize the speech rate difference between partners in a conversation (Manson et al., 2013;
54
55 Schultz, O'Brien, Phillips, & McFarland, 2016). Our results confirm previous evidence of this
56
57

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 phenomenon in a real communication setup (R. L. Street, Street, & Van Kleeck, 1983; R. L.
5
6 Street & Cappella, 1989) extending it to an interaction with a virtual non-adaptive partner and
7
8 with single words. Importantly, we demonstrate speech rate convergence in a population of
9
10 children with HL. While Freeman & Pisoni, (2017) found a less consistent convergence in
11
12 cochlear implant users compared to normal hearing children, this could be due to important
13
14 differences in the setup, which include using a natural and thus variable interlocutor (Freeman
15
16 & Pisoni) versus a constant speech rate in the present study, and the use of a sentence
17
18 repetition task versus a much easier object naming task. Thus, our experimental design, using
19
20 an alternating naming task, seems particularly suitable to assess convergence abilities even in
21
22 populations wherein these abilities may be impaired.
23
24
25

26 **Temporal accommodation abilities and speech-turn regularity**

27
28
29 The consistency of accent placement in regular turns for the two groups (NH-HL)
30
31 corroborates our previous results (Hidalgo et al., 2017). In that study, children had to name
32
33 words whose accent patterns (number of syllables) could be identical or not to those of the
34
35 virtual partner with which the children interacted; the turns of speech were thus more or less
36
37 regular. Overall, children were consistent in placing word accents, especially when the words
38
39 they had to produce had the same accent pattern as the words produced by the virtual partner.
40
41
42

43 In the present study, children in both groups (NH and HL) were more accurate during regular
44
45 than irregular turns in placing the stress of the word at the « expected » moment, that is
46
47 precisely in between two turns of the virtual partner. Interestingly, during regular turns,
48
49 children tended to place the stress slightly before this expected moment. This is reminiscent of
50
51 the negative asynchrony described in sensori-motor tapping tasks and has been interpreted in
52
53 terms of anticipatory behaviour (Repp, 2005) that can be absent in other species (Zarco,
54
55 Merchant, Prado, & Mendez, 2009; Patel, 2014; Merchant & Honing, 2014). By contrast,
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 during irregular turns, children were late in placing the word stress. During these (irregular)
4 turns, the distance between the consecutive accents of the PV was suddenly shortened, which
5 implied for the child a local readjustment of the internal temporal model of the alternation.
6
7 According to the Adaptation and Anticipation Model (ADAM) model (van der Steen &
8 Keller, 2013), interacting individuals generate temporal predictions on both their own and
9 their partner's actions. This relies on an anticipation mechanism based on the temporal
10 regularities of the partner's behavior and an adaptation mechanism allowing to adjust to local
11 temporal (phase) variations. During the irregular turns, children need both the anticipation
12 module to estimate the duration of the next interval and the adaptation module to make local
13 corrections to adjust to the temporal deviation of the VP.
14
15

16 ERP results indicate that the auditory system of normal-hearing children was sensitive to the
17 deviation from temporal expectations in speech turns, similarly to temporal deviations in
18 rhythmic musical sequences (Vuust et al., 2005). Interestingly, because the VP was following
19 a very slow pace (2000ms between words), temporal expectation possibly emerged from the
20 subdivision of this slow regular pace by the child turns (~1000ms). This is supported by the
21 correlation between the temporal accommodation capacities of the child and the amplitude of
22 the MMN-like effect. Interestingly, the MMN-like measure was issued from a temporal
23 cluster lasting up to ~500ms which may imply a role of attentive and conscious processing. In
24 other words, the better the child consciously perceives a change in the regularity of the
25 alternation, the better s/he can predict and adapt speech turns. This correlation of performance
26 of normal hearing children and children with hearing loss between perception and production
27 abilities suggests that anticipation and adaptation capacities during verbal interaction are
28 tightly related to the sensitivity of the auditory system in processing the temporal structure of
29 speech. The importance for speech accommodation of directing attention and adapting to
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 temporal deviations is particularly emphasized in deaf children after the rhythmic stimulation
4
5 session.
6
7
8
9
10
11
12
13

14 **Effect of the rhythmic training session on speech perception and temporal**
15 **accommodation abilities in children with HL**
16

17
18
19 Children did not show a greater convergence on virtual partner speech rate after the rhythmic
20 compared to the auditory training possibly because our design, with an artificial partner who
21 keeps the same speech rate along the blocks, renders the speech rate very easy to reproduce.
22
23
24
25

26
27 However, following only 30 minutes of rhythmic training, children with HL improved their
28 performance in the turn-taking task. More precisely, they became more accurate in temporally
29 placing the stress of the word following an irregular turn, wherein the virtual partner
30 anticipates the turn. This is the condition that requires the greatest flexibility to quickly adapt
31 to the anticipated turn. This result, coupled to the greater and longer-lasting effect on the
32 ERPs to deviant turns, suggests that the rhythmic training has engendered a more robust and
33 prompt processing of temporal deviations.
34
35
36
37
38
39
40
41
42

43 This is in line with previous findings showing that musicians are more consistent and accurate
44 than non-musicians in several types of temporal tasks, and are also more sensitive to temporal
45 deviations occurring in a regular auditory sequence (Drake, 1993; Rammsayer & Altenmüller,
46 2006; Repp, 2010; Yee, Holleran, & Riessjones, 1994; Jones & Yee, 1997). Moreover,
47 musical expertise improves temporal processing not only in music but also in speech at
48 different temporal scales (Cameron & Grahn, 2014; Geiser, Sandmann, Jäncke, & Meyer,
49 2010; Marie et al., 2011; Chobert J, Marie C, François C, Schön D, 2011; Elmer, Meyer, &
50
51
52
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 Jäncke, 2012; Sares, Foster, Allen, & Hyde, 2018). When we turn to populations with hearing
5
6 loss, cochlear implant users undergoing musical training show a greater sensitivity to fine
7
8 temporal modification of sounds, as reflected in larger ERP responses compared to cochlear
9
10 implant users without musical training (Timm et al., 2012). Moreover, children with cochlear
11
12 implant that practice regular singing show enhanced P3a-like responses to several sound
13
14 feature modifications compared to children with cochlear implant that do not practice regular
15
16 singing (Torppa, Huotilainen, Leminen, Lipsanen, & Tervaniemi, 2014). Similarly, in the
17
18 present study, music training mostly affected the amplitude of the MMN-like component,
19
20 which has been described as involved in attentional processes (Gray, Ambady, Lowenthal, &
21
22 Deldin, 2004). Thus, it seems that the rhythmic session has enhanced the children's ability to
23
24 make accurate temporal predictions on the occurrence of regular turns. Then, small deviations
25
26 from regularity may not only be automatically detected by the auditory system, but also
27
28 actively processed to adapt speech behavior. This effect is not visible for the same children
29
30 following 30 minutes of auditory training. This enhanced prediction is possibly mediated by a
31
32 greater involvement of the motor and premotor cortex in musicians (Krause, Schnitzler, &
33
34 Pollok, 2010) that in turn affects auditory perception (Merchant, Grahn, Trainor, Rohrmeier,
35
36 & Fitch, 2015). In other words, the greater implication of the motor system in auditory
37
38 perception improves the precision of temporal prediction of auditory events (Morillon,
39
40 Schroeder, & Wyart, 2014; Morillon & Baillet, 2017). Thus, active rhythmic training,
41
42 involving body movement and sensorimotor synchronization, has possibly increased the
43
44 involvement of motor system in temporal prediction, facilitating the detection of irregular
45
46 turns and temporal accommodation. Moreover, the active rhythmic training proposed here
47
48 stimulated not only the synchronization abilities on an external rhythm but also the
49
50 interpersonal coordination (for instance via rhythm repetition and rhythm exchanges). This
51
52 has possibly facilitated the realization of the picture naming alternation task as a joint action
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 to be realized in coordination with the partner. Interestingly, joint action planning compared
5
6 with individual action planning can result in amplitude modulations of the P300 (Kourtis et
7
8 al., 2013).
9

10
11 To conclude, while the regularity in speech turns in our paradigm was exaggerated compared
12
13 to natural conversation, one can imagine that the same mechanism is at work in both natural
14
15 and controlled exchanges. Thus, rhythmic training could be a powerful tool to facilitate
16
17 temporally accurate speech interaction, which in turn will improve speech coordination in
18
19 conversation (Konvalinka, Vuust, Roepstorff, & Frith, 2010), and boost engagement in
20
21 prosocial and cooperative attitudes (Cirelli, Wan, & Trainor, 2014; Valdesolo, Ouyang, &
22
23 DeSteno, 2010).
24
25

26 **Acknowledgments**

27
28
29
30 Research supported by grants of the Fondation Agir pour l'Audition (APA RD-2016-9),
31
32 ANR-16-CE28-0012 RALP, ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI)
33
34 and the Excellence Initiative of Aix-Marseille University (A*MIDEX). We wish to thank
35
36 Romane Pradels and Laura Leone for their contribution in speech data analysis.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Rhythmic training improves temporal conversational abilities in children with hearing loss

Table 1

Child	Age in months (years)	Duration of use of the device (in months)	Age at switch-on or HA wearing (in months)	Type of device	Current hearing Loss without HA(s) or CI(s)	Thresholds at 250-500-1000-2000 Hz	Onset of deafness
1	115 (9.7)	HA:51	64	HA	severe	10-25-35-25	unknown
2	94 (7.10)	CI :70	12 (HA) 24	CI	profound	30-30-30-30	congenital
3	92 (7.8)	CI:40 HA:47	52 18	CI + HA	profound	30-30-35-40	congenital
4	76 (6.6)	CI n°1:64 CI n°2: 64	12 12	+ CI	cophosis	35-25-25-25 35-25-25-25	perilingual
5	116 (9.8)	CI:80	30	CI	profound	30-25-25-20	congenital
6	82 (6.10)	HA:72	6	HA + HA	severe	20-20-25-35 25-30-30-35	congenital
7	85 (7.1)	HA:77	8	HA +	moderate	10-10-15-30 15-20 15-30	congenital

Rhythmic training improves temporal conversational abilities in children with hearing loss

			HA					
8	96 (8.8)	CI:66	16 (HA)	CI	profound	25-25-25-25		congenital
			30					
9	111(9.3)	CI n°1:104	7	CI		30-25-25-20		
		CI n°2:94	17	+	profound	30-30-35-35		unknown
				CI				
10	88 (7.4)	CI n°1:74	14	CI		25-35-40-35		
		CI n°2:65	23	+	profound	20-30-40-35		congenital
				CI				
11	123 (10.3)	HA:87	36	HA	severe	10-20-20-25		congenital
						15-20-25-25		
12	111 (9.3)	CI:93	18	CI	profound	35-25-40-30		congenital
				CI				
13	126 (10.6)	CI n°1:60	66	+	profound	15-20-20-35		
		CI n°2:23	103	CI		15-20-20-20		progressive
				CI				
14	93 (7.9)	CI n°1:60	23	+	profound	25-35-40-35		
		CI n°2:58	35	CI		20-30-40-35		congenital

Table 1. Demographic data of children with HL. HA: hearing aid; CI: cochlear implant (in case of bilateral implantation, n°1 means the duration of use of the 1st implant and n°2 the duration of use of the second contralateral implant, HA in parenthesis means the onset of

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 wearing a hearing device before implantation). Thresholds are given for each device (children
5
6 wearing a cochlear implant without contralateral hearing aid like Child n°2, did not have any
7
8 usable acoustic hearing in the unaided ear).
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3
4 Figures Legends

5
6 **Figure 1. Task and procedure.** A. Schematic representation of the task. The virtual partner
7 pronounces isolated words at *regular* (blue, 2000 ms) or *irregular* (red, 1600 ms) intervals.
8 The child names the objects presented on the screen in alternation with the virtual partner.
9 The speech rate of the virtual partner can be *fast* (light blue, top panel) or *slow* (dark blue,
10 bottom panel). The vertical arrows indicate the p-center placement of the words. IWI stands
11 for Inter Words Intervals, namely the distance between the virtual partner p-center and the
12 child p-center. WD stands for Words Duration, namely the duration of the words pronounced
13 by the child. B. Schematic representation of the procedure. Normal hearing (NH) children ran
14 the task once and children with hearing loss (HI) ran the task twice (order was
15 counterbalanced). For all figures, the ear represents children with normal hearing, the maracas
16 represent children with hearing loss after rhythmic training, and the loudspeaker represents
17 children with hearing loss after auditory training.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 **Figure 2. Behavioral results for NH children.** A. Convergence results. Effect of the virtual
36 partner's speech rate on mean word duration in the fast and slow conditions. B.
37 Accommodation results. Effect of the regularity of alternation on children's stress placement.
38 Left panel depicts the consistency of stress placement, represented by the vector length (from
39 0 to 1). Higher values represent better performances. Right panel depicts the accuracy of
40 stress placement. Time zero represents a stress placement at the expected time. Positive values
41 represent late responses, and negative values represent early responses.
42
43
44
45
46
47
48
49
50
51
52

53 **Figure 3. Behavioral results for HI children.** A. Convergence results following a 30-min
54 rhythmic training. Effect of the virtual partner speech rate on mean word duration in the fast
55
56
57

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3 and slow conditions. B. Accommodation results following a 30-min rhythmic training. Effect
4 of the regularity of alternation on children stress placement. Left panel depicts the consistency
5 of stress placement, represented by the vector length (from 0 to 1). Higher values represent
6 better performances. Right panel depicts the accuracy of stress placement. Time zero
7 represents a stress placement at the expected time. Positive values represent late responses,
8 and negative values represent early responses. C-D. Convergence and accommodation results
9 following a 30-min auditory training.
10
11
12
13
14
15
16
17
18
19
20

21 **Figure 4. Electrophysiological (EEG) results.** Grand average of Event Related Potentials for
22 regular and irregular trials. Time zero is the onset of the word pronounced by the virtual
23 partner. Shadowed areas represent significant time clusters ($p < .05$). A. NH children. B. HI
24 children after rhythmic training. C. HI children after auditory training.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3

4 **References**

5
6 Bates, D., & Maechler, M. (2009). *Dai B lme4: Linear mixed-effects models using S4 classes.*

7
8 R package version 0.999375-28.
9

10
11 Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally Regular

12
13 Musical Primes Facilitate Subsequent Syntax Processing in Children with Specific

14
15 Language Impairment. *Frontiers in Neuroscience, 10*, 245.

16
17 <https://doi.org/10.3389/fnins.2016.00245>
18
19

20
21 Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical*

22
23 *Software, 31*(10), 1–21. <https://doi.org/10.1002/wics.10>
24
25

26
27 Bigi, B. (2015). SPPAS-multi-lingual approaches to the automatic annotation of speech. *The*

28
29 *Phonetician-International Society of Phonetic Sciences*, (111-112), 54-69.
30

31
32 Cameron, D. J., & Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize

33
34 to rhythms without a musical beat. *Frontiers in Human Neuroscience, 8*(December),

35
36 1003. <https://doi.org/10.3389/fnhum.2014.01003>
37
38

39
40 Cannard, C., Blaye, A., Scheuner, N., & Bonthoux, F. (2005). Picture naming in 3- to 8-year-

41
42 old French children: methodological considerations for name agreement. *Behavior*

43
44 *Research Methods, 37*(3), 417–425. <https://doi.org/10.3758/BF03192710>
45

46
47 Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: rhythmic

48
49 priming and audio-motor training affect speech perception. *Acta Psychologica, 155*, 43–

50
51 50. <https://doi.org/10.1016/j.actpsy.2014.12.002>
52

53
54 Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming

55
56 enhances speech production abilities: Evidence from prelingually deaf children.
57
58
59
60

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 *Neuropsychology*, 29(1), 102–107. <https://doi.org/10.1037/neu0000115>
5
6 Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of
7
8 speech. *Neuropsychologia*, 50(11), 2652–8.
9
10 <https://doi.org/10.1016/j.neuropsychologia.2012.07.018>
11
12
13 Chern, A., Tillmann, B., Vaughan, C., Gordon, R. L., & Gordon, R. (2017). New evidence of
14
15 a rhythmic priming effect that enhances grammaticality judgments in children.
16
17 <https://doi.org/10.1101/193961>
18
19
20 Chobert J, Marie C, François C, Schön D, B. M. (2011). Enhanced passive and active
21
22 processing of syllables in musician children. *J Cogn Neurosc.*, 23(12), 3874–3887.
23
24 https://doi.org/10.1162/jocn_a_00088
25
26
27 Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2014). Fourteen-month-old infants use
28
29 interpersonal synchrony as a cue to direct helpfulness. *Philosophical Transactions of the*
30
31 *Royal Society B: Biological Sciences*, 369(1658), 20130400–20130400.
32
33 <https://doi.org/10.1098/rstb.2013.0400>
34
35
36 Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of*
37
38 *Phonetics*, 26(2), 145–171. <https://doi.org/10.1006/jpho.1998.0070>
39
40
41
42 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-
43
44 trial EEG dynamics including independent component analysis. *Journal of Neuroscience*
45
46 *Methods*, 134, 9–21. Retrieved from
47
48 <https://cloudfront.escholarship.org/dist/prd/content/qt52k1t4sz/qt52k1t4sz.pdf>
49
50
51
52 Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult
53
54 nonmusicians. *Perception & Psychophysics*, 53(1), 25–33.
55
56
57
58
59
60

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 <https://doi.org/10.3758/BF03211712>
5
6 Elmer, S., Meyer, M., & Jäncke, L. (2012). The spatiotemporal characteristics of elementary
7
8 audiovisual speech and music processing in musically untrained subjects. *International*
9
10 *Journal of Psychophysiology*, 83(3), 259–268.
11
12 <https://doi.org/10.1016/j.ijpsycho.2011.09.011>
13
14
15 Falk, S., Müller, T., & Dalla Bella, S. (2015). Non-verbal sensorimotor timing deficits in
16
17 children and adolescents who stutter. *Frontiers in Psychology*, 6, 847.
18
19 <https://doi.org/10.3389/fpsyg.2015.00847>
20
21
22 Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music
23
24 Training Increases Phonological Awareness and Reading Skills in Developmental
25
26 Dyslexia: A Randomized Control Trial. *Plos One*, 10(9), e0138715.
27
28 <https://doi.org/10.1371/journal.pone.0138715>
29
30
31
32 Freeman, V., & Pisoni, D. B. (2017). Speech rate, rate-matching, and intelligibility in early-
33
34 implanted cochlear implant users. *The Journal of the Acoustical Society of America*,
35
36 142(2), 1043–1054. <https://doi.org/10.1121/1.4998590>
37
38
39 Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception -
40
41 training increases hierarchical metre processing. *European Journal of Neuroscience*,
42
43 32(11), 1979–1985. <https://doi.org/10.1111/j.1460-9568.2010.07462.x>
44
45
46 Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of
47
48 attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, 40(2),
49
50 216–224. [https://doi.org/10.1016/S0022-1031\(03\)00092-1](https://doi.org/10.1016/S0022-1031(03)00092-1)
51
52
53 Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3 critical review. *Neuroscience & Biobehavioral Reviews*, 86(December), 150–165.

4
5
6 <https://doi.org/10.1016/j.neubiorev.2017.12.002>

7
8 Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic

9
10 training on children with hearing loss. *Hearing Research*, 351(May), 11–18.

11
12
13 <https://doi.org/10.1016/j.heares.2017.05.006>

14
15 Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the

16
17 timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in*

18
19
20 *Psychology*, 6(September), 1–12. <https://doi.org/10.3389/fpsyg.2015.01492>

21
22
23 Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of

24
25 speech rhythm during joint story building. *Frontiers in Psychology*, 6(June), 1–6.

26
27
28 <https://doi.org/10.3389/fpsyg.2015.00797>

29
30 Hoffman, M. F., Quittner, a. L., & Cejas, I. (2014). Comparisons of Social Competence in

31
32 Young Children With and Without Hearing Loss: A Dynamic Systems Framework.

33
34
35 *Journal of Deaf Studies and Deaf Education*, 20(2), 115–124.

36
37
38 <https://doi.org/10.1093/deafed/enu040>

39
40 Holt, C. M., Demuth, K., & Yuen, I. (2016). The Use of Prosodic Cues in Sentence

41
42 Processing by Prelingually Deaf Users of Cochlear Implants. *Ear and Hearing*, 37(4),

43
44
45 e256–e262. <https://doi.org/10.1097/AUD.0000000000000253>

46
47 Holt, C. M., Yuen, I., & Demuth, K. (2017). Discourse Strategies and the Production of

48
49
50 Prosody by Prelingually Deaf Adolescent Cochlear Implant Users, 101–108.

51
52 Jones, M. R., & Yee, W. (1997). Sensitivity to time change: The role of context and skill.

53
54
55 *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 693–

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 709. <https://doi.org/10.1037/0096-1523.23.3.693>
5
6 Klein, C., Liem, F., Hänggi, J., Elmer, S., & Jäncke, L. (2015). The “silent” imprint of
7
8 musical training. *Human Brain Mapping, 00*(August), n/a-n/a.
9
10 <https://doi.org/10.1002/hbm.23045>
11
12
13 Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me:
14
15 Continuous mutual prediction and adaptation in joint tapping. *The Quarterly Journal of*
16
17 *Experimental Psychology, 63*(11), 2220–2230.
18
19 <https://doi.org/10.1080/17470218.2010.497843>
20
21
22
23 Kotz, S. A., Gunter, T. C., & Wonneberger, S. (2005). The basal ganglia are receptive to
24
25 rhythmic compensation during auditory syntactic processing: ERP patient data.
26
27 <https://doi.org/10.1016/j.bandl.2005.07.039>
28
29
30 Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other people’s
31
32 actions in joint action planning: An EEG study. *Social Neuroscience, 8*(1), 31–42.
33
34 <https://doi.org/10.1080/17470919.2012.694823>
35
36
37
38 Krause, V., Schnitzler, A., & Pollok, B. (2010). Functional network interactions during
39
40 sensorimotor synchronization in musicians and non-musicians. *NeuroImage, 52*, 245–
41
42 251. <https://doi.org/10.1016/j.neuroimage.2010.03.081>
43
44
45 Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2014). lmerTest: tests in linear
46
47 mixed effects models. R package version 2.0–20. Available at [http://CRAN.R-](http://CRAN.R-project.org/package=lmerTest)
48
49 [project.org/package=lmerTest](http://CRAN.R-project.org/package=lmerTest)
50
51
52 Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for
53
54 processing models of language. *Frontiers in Psychology, 6*(JUN), 1–17.
55
56
57
58
59
60

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 <https://doi.org/10.3389/fpsyg.2015.00731>
5
- 6 Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical
7
8 aptitude: ERPs and individual differences. *Brain and Language*.
9
10 <https://doi.org/10.1016/j.bandl.2016.01.001>
11
12
- 13 Manson, J. H., Bryant, G. a., Gervais, M. M., & Kline, M. a. (2013). Convergence of speech
14
15 rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34, 419–426.
16
17 <https://doi.org/10.1016/j.evolhumbehav.2013.08.001>
18
19
- 20 Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words.
21
22 *Journal of Cognitive Neuroscience*, 23(2), 294–305.
23
24 <https://doi.org/10.1162/jocn.2010.21413>
25
26
- 27 Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat:
28
29 a neural perspective across humans and non-human primates. *Philosophical*
30
31 *Transactions of the Royal Society of London. Series B, Biological Sciences*, 370,
32
33 20140093. <https://doi.org/10.1098/rstb.2014.0093>
34
35
36
- 37 Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic
38
39 entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in*
40
41 *Neuroscience*, 7(8 JAN), 1–8. <https://doi.org/10.3389/fnins.2014.00274>
42
43
44
- 45 Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention.
46
47 *Proceedings of the National Academy of Sciences of the United States of America*,
48
49 114(42), E8913–E8921. <https://doi.org/10.1073/pnas.1705373114>
50
51
- 52 Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal
53
54 precision of auditory attention. *Nature Communications*, 5, 5255.
55
56
57
58
59
60

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 <https://doi.org/10.1038/ncomms6255>
5
- 6 Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between
7
8 early rhythm skills, musical training, and phonological awareness. *Reading and Writing*,
9
10 26(5), 739–769. <https://doi.org/10.1007/s11145-012-9389-0>
11
12
- 13 Morton, J., Marcus, S. M., & Frankish, C. R. (1976). Perceptual Centers (P-centers).
14
15 *Psychological Review*, 83(5), 405–408. <https://doi.org/10.1037/0033-295X.83.5.405>
16
17
- 18 Pallone, G., Boussard, P., Daudet, L., Guillemain, P., & Kronland-Martinet, R. (1999,
19
20 December). A wavelet based method for audio-video synchronization in broadcasting
21
22 applications. In *Proceedings of the International Conference on Digital Audio Effects*,
23
24 Norway.
25
26
- 27 Patel, A. D. (2014). The Evolutionary Biology of Musical Rhythm: Was Darwin Wrong?
28
29 *PLoS Biology*, 12(3). <https://doi.org/10.1371/journal.pbio.1001821>
30
31
32
- 33 Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to
34
35 Comprehension. *Frontiers in Psychology*, 3(September), 320.
36
37 <https://doi.org/10.3389/fpsyg.2012.00320>
38
39
- 40 Petersen, B., Weed, E., Sandmann, P., Brattico, E., Hansen, M., Sørensen, S. D., & Vuust, P.
41
42 (2015). Brain Responses to Musical Feature Changes in Adolescent Cochlear Implant
43
44 Users. *Frontiers in Human Neuroscience*, 9, 7.
45
46 <https://doi.org/10.3389/fnhum.2015.00007>
47
48
- 49 Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and
50
51 comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–47.
52
53 <https://doi.org/10.1017/S0140525X12001495>
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3
4 Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, 31(3–4), 599–611.

5
6 <https://doi.org/10.1016/J.WOCN.2003.08.001>

7
8
9 Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., ... Tillmann,
10
11 B. (2013). Rhythmic auditory stimulation influences syntactic processing in children
12
13 with developmental language disorders. *Neuropsychology*, 27(1), 121–31.

14
15 <https://doi.org/10.1037/a0031277>

16
17
18 Rammsayer, T., & Altenmüller, E. (2006). Temporal Information Processing in Musicians
19
20 and Nonmusicians. *Music Perception*, 24(1), 37–48.

21
22 <https://doi.org/10.1525/mp.2006.24.1.37>

23
24
25 Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature.

26
27 *Psychonomic Bulletin & Review*, 12(6), 969–992. <https://doi.org/10.3758/BF03206433>

28
29
30 Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music
31
32 training and task experience. *Human Movement Science*, 29(2), 200–213.

33
34 <https://doi.org/10.1016/j.humov.2009.08.002>

35
36
37 Roy, A. T., Scattergood-Keeper, L., Carver, C., Jiradejvong, P., Butler, C., & Limb, C. J.
38
39 (2014). Evaluation of a Test Battery to Assess Perception of Music in Children With
40
41 Cochlear Implants. *JAMA Otolaryngology-- Head & Neck Surgery*, 140(6), 1–8.

42
43 <https://doi.org/10.1001/jamaoto.2014.341>

44
45
46
47 Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and Time Processing in
48
49 Speech and Tones: The Effects of Musical Training and Attention. *Journal of Speech*

50
51 *Language and Hearing Research*, 61(3), 496. https://doi.org/10.1044/2017_JSLHR-S-

52
53
54 17-0207

- 1 Rhythmic training improves temporal conversational abilities in children with hearing loss
2
3
4 Schön, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: perspectives
5
6 for language rehabilitation. *Annals of the New York Academy of Sciences*, *1337*(1), 32–
7
8 39. <https://doi.org/10.1111/nyas.12635>
9
10
11 Schultz, B. G., O'Brien, I., Phillips, N., & McFarland, D. H. (2016). Speech rates converge in
12
13 scripted turn-taking conversations. *Applied Psycholinguistics*, *37*, 1201–1220.
14
15 <https://doi.org/10.1017/S0142716415000545>
16
17
18 Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where.
19
20 *Topics in Cognitive Science*, *1*(2), 353–367. <https://doi.org/10.1111/j.1756->
21
22 [8765.2009.01024.x](https://doi.org/10.1111/j.1756-8765.2009.01024.x)
23
24
25 Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., & Vatovec, J. (2012). The music
26
27 perception abilities of prelingually deaf children with cochlear implants. *International*
28
29 *Journal of Pediatric Otorhinolaryngology*, *76*(10), 1392–400.
30
31 <https://doi.org/10.1016/j.ijporl.2012.07.004>
32
33
34
35 Street, R. (1984). Speech convergence and speech evaluation in fact-finding interviews.
36
37 *Human Communication Research*, *11*(2), 139–169. <https://doi.org/10.1111/j.1468->
38
39 [2958.1984.tb00043.x](https://doi.org/10.1111/j.1468-2958.1984.tb00043.x)
40
41
42 Street, R. L., & Cappella, J. N. (1989). Social and linguistic factors influencing adaptation in
43
44 children's speech. *Journal of Psycholinguistic Research*, *18*(5), 497–519.
45
46 <https://doi.org/10.1007/BF01067313>
47
48
49 Street, R. L., Street, N. J., & Van Kleek, A. (1983). Speech convergence among talkative and
50
51 reticent three year-olds. *Language Sciences*, *5*, 79–96. <https://doi.org/10.1016/S0388->
52
53 [0001\(83\)80015-1](https://doi.org/10.1016/S0388-0001(83)80015-1)
54
55
56
57
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3
4 Timm, L., Agrawal, D., C. Viola, F., Sandmann, P., Debener, S., Büchner, A., ... Wittfoth, M.
5
6 (2012). Temporal Feature Perception in Cochlear Implant Users. *PLoS ONE*, 7(9),
7
8 e45375. <https://doi.org/10.1371/journal.pone.0045375>
9

10
11 Torppa, R., Huotilainen, M., Leminen, M., Lipsanen, J., & Tervaniemi, M. (2014). Interplay
12
13 between singing and cortical processing of music : A longitudinal study in children with
14
15 cochlear implants Interplay between singing and cortical processing of music : A
16
17 longitudinal study in children with cochlear implants. *Frontiers in Psychology*,
18
19 5(December), 1–16. <https://doi.org/10.3389/fpsyg.2014.01389>
20
21

22
23 Torppa, R., Salo, E., Makkonen, T., Loimo, H., Pykäläinen, J., Lipsanen, J., ... Huotilainen,
24
25 M. (2012). Cortical processing of musical sounds in children with Cochlear Implants.
26
27 *Clinical Neurophysiology*, 123(10), 1966–1979.
28
29 <https://doi.org/10.1016/j.clinph.2012.03.008>
30
31

32
33 Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony
34
35 promotes cooperative ability. *Journal of Experimental Social Psychology*, 46(4), 693–
36
37 695. <https://doi.org/10.1016/j.jesp.2010.03.004>
38

39
40 van der Steen, M. C. M., & Keller, P. E. (2013). The ADaptation and Anticipation Model
41
42 (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7(June),
43
44 253. <https://doi.org/10.3389/fnhum.2013.00253>
45

46
47 Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., &
48
49 Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive neuronal
50
51 responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, 24(2),
52
53 560–4. <https://doi.org/10.1016/j.neuroimage.2004.08.039>
54
55

56
57 Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking.
58
59
60

1 Rhythmic training improves temporal conversational abilities in children with hearing loss

2
3 *Psychonomic Bulletin & Review*, 12(6), 957–68. Retrieved from

4 <http://www.ncbi.nlm.nih.gov/pubmed/16615316>

5
6
7
8 Wong, C. L., Ching, T. Y. C., Cupples, L., Button, L., Leigh, G., Marnane, V., ... Martin, L.

9 (2017). Psychosocial Development in 5-Year-Old Children With Hearing Loss Using

10 Hearing Aids or Cochlear Implants. *Trends in Hearing*, 21, 233121651771037.

11 <https://doi.org/10.1177/2331216517710373>

12
13
14
15
16
17 Yee, W., Holleran, S., & Riessjones, M. (1994). Sensitivity to event timing in regular and

18 irregular sequences: Influences of musical skill. *Perception & Psychophysics*, 56(4),

19 461–471. Retrieved from <https://link.springer.com/content/pdf/10.3758/BF03206737.pdf>

20
21
22
23
24
25 Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond Timing in Primates:

26 Comparison of Interval Production Between Human Subjects and Rhesus Monkeys.

27
28
29
30 *Journal of Neurophysiology*, 102(6), 3191–3202. <https://doi.org/10.1152/jn.00066.2009>

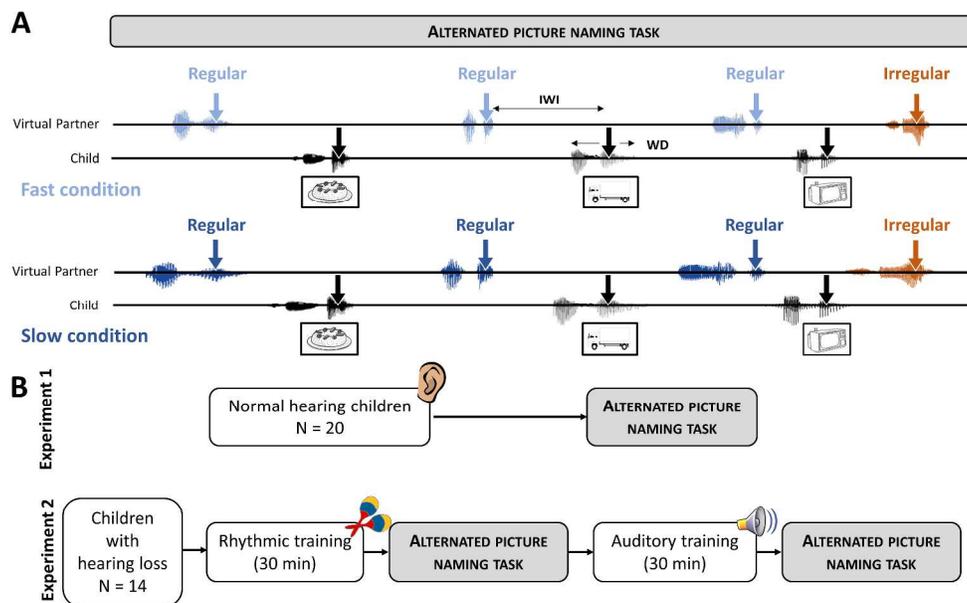


Figure 1. Task and procedure. A. Schematic representation of the task. The virtual partner pronounces isolated words at *regular* (blue, 2000 ms) or *irregular* (red, 1600 ms) intervals. The child names the objects presented on the screen in alternation with the virtual partner. The speech rate of the virtual partner can be *fast* (light blue, top panel) or *slow* (dark blue, bottom panel). The vertical arrows indicate the p-center placement of the words. IWI stands for Inter Words Intervals, namely the distance between the virtual partner p-center and the child p-center. WD stands for Words Duration, namely the duration of the words pronounced by the child. B. Schematic representation of the procedure. Normal hearing (NH) children ran the task once and children with hearing loss (HI) ran the task twice (order was counterbalanced). For all figures, the ear represents children with normal hearing, the maracas represent children with hearing loss after rhythmic training, and the loudspeaker represents children with hearing loss after auditory training.

340x209mm (300 x 300 DPI)

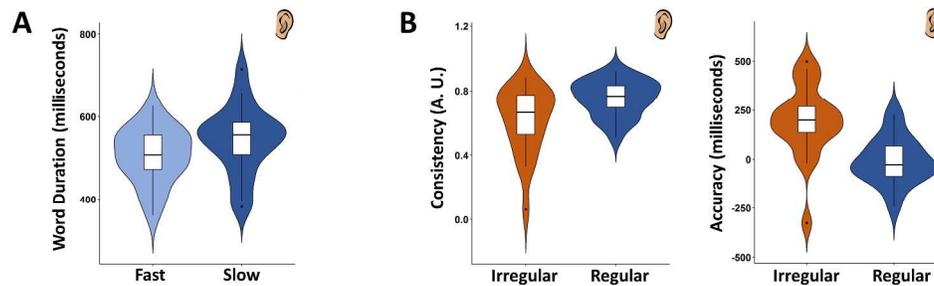


Figure 2. Behavioral results for NH children. A. Convergence results. Effect of the virtual partner's speech rate on mean word duration in the fast and slow conditions. B. Accommodation results. Effect of the regularity of alternation on children's stress placement. Left panel depicts the consistency of stress placement, represented by the vector length (from 0 to 1). Higher values represent better performances. Right panel depicts the accuracy of stress placement. Time zero represents a stress placement at the expected time. Positive values represent late responses, and negative values represent early responses.

340x209mm (300 x 300 DPI)

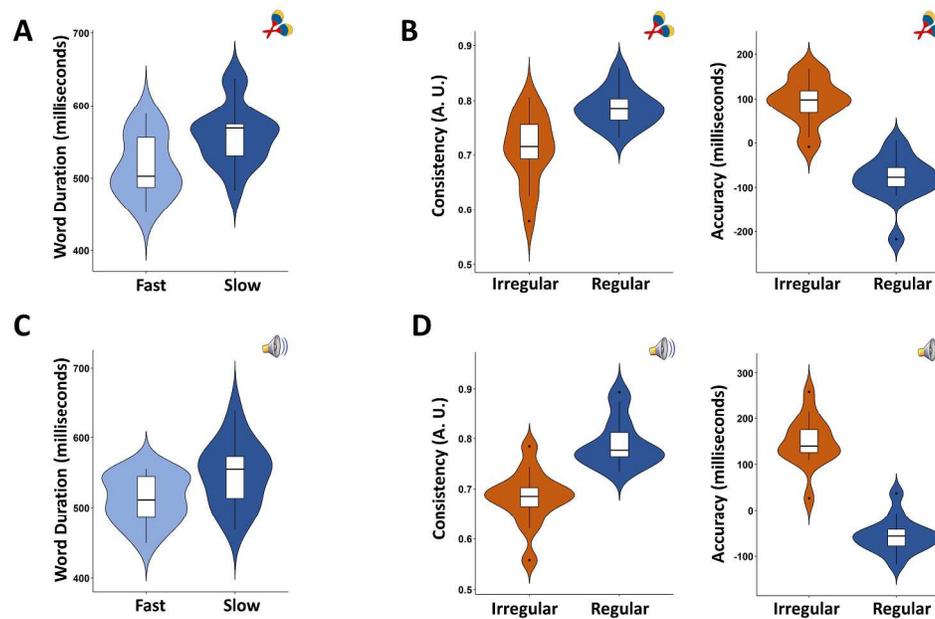


Figure 3. Behavioral results for HI children. A. Convergence results following a 30-min rhythmic training. Effect of the virtual partner speech rate on mean word duration in the fast and slow conditions. B. Accommodation results following a 30-min rhythmic training. Effect of the regularity of alternation on children stress placement. Left panel depicts the consistency of stress placement, represented by the vector length (from 0 to 1). Higher values represent better performances. Right panel depicts the accuracy of stress placement. Time zero represents a stress placement at the expected time. Positive values represent late responses, and negative values represent early responses. C-D. Convergence and accommodation results following a 30-min auditory training.

340x209mm (300 x 300 DPI)

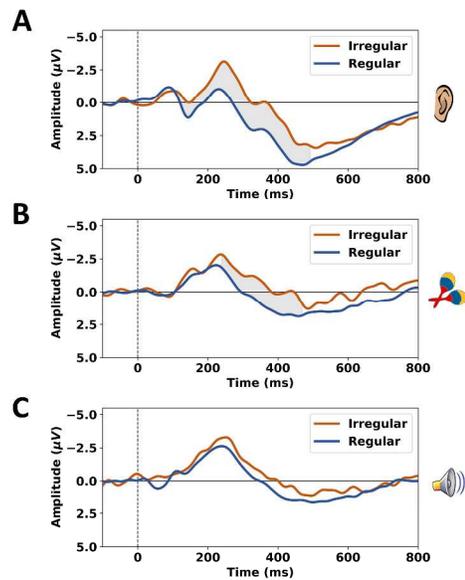


Figure 4. Electrophysiological (EEG) results. Grand average of Event Related Potentials for regular and irregular trials. Time zero is the onset of the word pronounced by the virtual partner. Shaded areas represent significant time clusters ($p < .05$). A. NH children. B. HI children after rhythmic training. C. HI children after auditory training.

340x209mm (300 x 300 DPI)

Partie Discussion

5. Rappel des résultats

Au cours de cette thèse, nous avons mené trois expériences associant des mesures comportementales et électrophysiologiques afin d'analyser les capacités de perception et de production rythmiques musicales et linguistiques des enfants sourds implantés cochléaires et/ou appareillés en situation d'interaction. Nous avons également testé les effets d'une stimulation rythmique d'une durée de 30 minutes sur ces capacités rythmiques impliquant le développement de prédictions temporelles.

Les résultats de ces trois expériences, nous ont permis d'affiner notre compréhension des capacités de traitement temporel chez l'enfant sourd et de faire le lien avec leurs difficultés linguistiques à des niveaux de langage plus intégrés comme l'interaction verbale, comparé à ceux habituellement testés comme la perception de la parole en situation isolée.

Dans **l'expérience n°1**, nous avons testé la capacité des enfants Normo-Entendants (NE) et Enfants Sourds (ES) à s'accommoder temporellement à la vitesse d'une alternance (lente versus rapide) de parole imposée par un partenaire virtuel ainsi que leur capacité à demeurer consistants dans cette alternance régulière en fonction de la métrique (identique versus différente) des mots échangés entre les deux partenaires. Nous avons aussi mesuré les effets d'un entraînement rythmique versus linguistique sur ces capacités temporelles chez l'enfant sourd.

Les résultats de cette 1ère expérience ont montré que les ES comme les enfants NE perçoivent et s'accommodent spontanément à la vitesse d'un échange de paroles. En fonction de la vitesse donnée par le partenaire virtuel, les enfants des deux groupes réduisent ou allongent la durée des intervalles qui séparent leurs productions verbales. Cependant, contrairement aux enfants NE qui sont plus réguliers lorsque les paires de mots échangées contiennent des patterns accentuels identiques, les ES ne sont pas sensibles à la régularité de la métrique des mots échangés. Ce n'est qu'après la séance de stimulation rythmique que les ES deviennent sensibles, comme les enfants NE, à la régularité des patterns accentuels des mots.

Dans **l'expérience n°2**, nous avons cherché à savoir si les ES, comme les enfants NE, étaient capables d'imiter la durée des mots du partenaire virtuel autrement dit s'ils pouvaient converger vers son débit de parole (lent versus rapide) et s'ils étaient capables de s'adapter à des variations temporelles de la régularité de l'alternance de parole (sans modification de la métrique contenue dans les mots échangés). Nous avons complété ces mesures comportementales par des mesures électrophysiologiques pour analyser par une mesure de la négativité de discordance (MMN), leur sensibilité aux variations temporelles séparant les productions du partenaire virtuel (intervalles réguliers versus irréguliers). Enfin, nous avons mesuré les effets d'un entraînement rythmique versus auditif sur ces capacités de production et de perception temporelles de la parole en interaction chez l'enfant sourd. Nous avons également analysé le lien entre les capacités de perception et de production de la parole des deux groupes d'enfants et leurs capacités rythmiques simples de type sensori-moteur.

Les résultats de la deuxième expérience ont mis en évidence des capacités de discrimination et d'imitation de la durée des mots du partenaire virtuel similaires chez les ES et les enfants NE. Autrement dit, les ES ont montré des capacités de convergence sur le débit de parole du partenaire : ils raccourcissent ou allongent la durée des mots qu'ils prononcent lorsque le partenaire augmente ou réduit son débit de parole. Ils ont également manifesté, comme les enfants NH, une meilleure capacité d'adaptation lorsque le partenaire virtuel est régulier plutôt qu'irrégulier : les enfants des deux groupes présentent une meilleure consistance et précision de leurs productions lors des intervalles réguliers. Les résultats électrophysiologiques ont par ailleurs révélé une différence de traitement perceptif des déviations temporelles entre les enfants NE et les ES: après la stimulation auditive, même si les ES perçoivent une différence entre les intervalles réguliers et les intervalles irréguliers, ils ne la perçoivent pas avec la même acuité que les enfants NE (MMN-like) alors qu'après la séance de stimulation rythmique, leur perception de la déviation temporelle va se prolonger comme chez les enfants NE. En outre, après la séance de stimulation rythmique, les ES améliorent la précision temporelle de leur dénomination lors des intervalles irréguliers et on retrouve une corrélation entre la qualité de la perception temporelle (mesurée par la durée de la MMN) et la précision du placement des P-centers sur les intervalles irréguliers chez les ES comme chez les enfants NE.

L'expérience n°3 visait à analyser de manière plus précise, les capacités et déficits de synchronisation rythmique des enfants sourds sur des stimuli avec une complexité acoustique simple (i.e. métronome) et complexe (i.e. musique) mais également avec une structure temporelle simple (i.e. pulsation) et complexe (i.e. patterns rythmiques). Ces stimuli rythmiques faisant appel à différents types de traitement temporel, nous avons cherché à mettre en lien les déficits rythmiques sur ces différentes tâches avec leurs déficits linguistiques en répétition de phrases en tenant compte de leurs capacités mnésiques. Ces résultats ont été comparés à ceux d'une population contrôle d'enfants normo-entendants.

Les résultats de cette troisième expérience montrent que les ES ont des performances de synchronisation inférieures à celles des enfants NE sur les 3 types de stimuli (i.e. tapping sur métronome, tapping sur extraits musicaux et tapping sur patterns rythmiques de complexité croissante), mais que la différence de performances entre les enfants NE et les ES est plus importante dans la tâche de reproduction de patterns rythmiques pour laquelle les stimuli sont simples d'un point de vue acoustique mais qui comportent la structure temporelle la plus complexe des trois tâches. Voir la figure 5.1 pour un récapitulatif des trois expériences et des résultats obtenus.

Expérience	Objectif	Mesure	Résultats
1	<p>a) Analyser la capacité des enfants NE (■) et ES (■ ■) à:</p> <ul style="list-style-type: none"> s'adapter à la vitesse d'une alternance de parole avec un partenaire virtuel imiter la régularité d'une alternance de parole en fonction de la régularité métrique des mots échangés avec un partenaire <p>b) Analyser les effets d'une stimulation rythmique ■ versus linguistique ■ sur ces 2 capacités</p>	<p>Durée des intervalles entre les tours ➤ tours lents versus rapide</p> <p>Consistance (longueur vectorielle) des productions ➤ métrique des échanges régulière versus irrégulière</p>	<p>✓ ✓ ✓ effet de la durée des tours</p> <p>✓ ✗ ✓ effet de la métrique des mots</p>  <p>➤ de la sensibilité à la régularité de la métrique des mots échangés</p>
2	<p>a) Analyser la capacité des enfants NE (■) et ES (■ ■) à:</p> <ul style="list-style-type: none"> converger vers le débit de parole d'un partenaire virtuel lors d'un échange de paroles détecter une violation de la régularité de la parole d'un partenaire s'adapter à la régularité et à l'irrégularité de la parole d'un partenaire <p>b) Analyser les effets d'une stimulation rythmique ■ versus auditive ■ sur ces 3 capacités</p>	<p>Durée des mots ➤ débit de parole lent versus rapide</p> <p>MMN & MMN-like ➤ tours réguliers versus irréguliers</p> <p>Consistance (longueur vectorielle) & Précision (angle) ➤ tours réguliers versus irréguliers</p>	<p>✓ ✓ ✓ effet du débit de parole</p> <p>✓ ✓ ✓ MMN effet de la violation de la régularité des tours</p> <p>✓ ✗ ✓ MMN-like (239 - 469 ms) effet de la violation de la régularité des tours</p> <p>✓ ✓ ✓ effet de la régularité</p>  <p>➤ sensibilité à l'irrégularité des tours</p> <p>➤ précision des productions dans les tours irréguliers</p>
3	<p>Analyser les compétences rythmiques (synchronisation sensori-motrice) des ES comparé aux enfants NE:</p> <ul style="list-style-type: none"> Sur des stimuli acoustiques simples et complexes Sur des séquences temporelles simples et complexes 	<p>Consistance du tapping ➤ métronome ➤ extraits musicaux</p> <p>Scores de synchronisation ➤ séquences de rythmes complexes</p>	<p>NE ➤ ES</p> <p>NE ➤ ES</p> <p>NE ➤ ES</p>

■ Enfants normo-entendants

■ Enfants sourds après stimulation rythmique

■ Enfants sourds après stimulation auditive

Figure 5.1. Récapitulatif des trois expériences et des résultats obtenus

D'une manière générale, les résultats de ces trois études chez les enfants sourds nous ont permis de distinguer des capacités et déficits de perception temporelle qui impliquent des traitements différents et les effets d'une stimulation rythmique sur leurs capacités linguistiques de nature temporelle. Dans la première partie de cette discussion, nous discuterons des liens entre les différents déficits rythmiques des enfants sourds et les altérations de processus cognitifs nécessaires au traitement du langage. Dans la deuxième partie, nous discuterons des effets de la stimulation rythmique sur leurs compétences temporelles en perception et production de la parole en contexte d'interaction dans le cadre du modèle ADAM. Dans la troisième partie, nous discuterons dans le cadre de la surdité, des différentes théories qui proposent des liens entre stimulation rythmique de type musical et amélioration des compétences langagières à différentes échelles temporelles.

1. Y-a-t-il une problématique de traitement temporel chez l'enfant sourd et si oui, comment la caractériser ?

Les différentes habiletés rythmiques (linguistiques et musicales) testées dans les expériences n°1, 2 et 3 nous ont permis d'avoir une vision plus précise des capacités temporelles des ES, pour la majorité d'entre eux implantés cochléaires, et de faire des hypothèses quant à la nature des processus de traitement temporels possiblement altérés.

a. Préservation de la capacité de discrimination/reproduction relative de durées dans la parole

Les résultats de l'expérience n°1 et n°2 montrent que les ES ont une perception préservée de la durée des événements de parole.

Dans l'expérience n°1, les ES comme les enfants NE, adaptent spontanément la durée des tours de parole en fonction de la vitesse de l'alternance imposée par un partenaire. Ce résultat semble signifier que dans le cadre d'une interaction verbale, la discrimination d'intervalles temporels variant de 2600 ms à 3200 ms (soit une différence de l'ordre de 600 ms), est une capacité préservée chez les ES et que ces derniers sont également capables d'adapter spontanément la planification de leur production de parole afin d'imiter de manière relative la durée des intervalles imposée par un partenaire.

La mesure de la convergence dans l'expérience n°2, nous a permis d'aller plus loin dans l'analyse de la perception de la durée d'un matériel verbal et nous montre que les ES ont également la capacité de discriminer la longueur des mots et qu'ils imitent spontanément ces différences de durées dans le cadre d'une interaction verbale. Ce dernier résultat va à l'encontre des résultats retrouvés dans l'étude de Freeman et collaborateurs, (2017) : dans cette étude, les ES présentaient un déficit de convergence sur le débit de parole comparé aux enfants NE (Freeman & Pisoni, 2017). Cependant, le protocole expérimental de Freeman et collaborateurs (2017) a été élaboré à partir de conversations avec un partenaire réel. La mesure des capacités de convergence a ainsi été réalisée à partir d'un stimulus de parole naturelle avec un débit beaucoup plus variable que celui utilisé dans notre expérience. On peut penser que le protocole de l'expérience n°2, par la régularité du débit de parole et de la structure métrique des mots à travers les essais, a favorisé, contrairement au protocole de

Freeman et collaborateurs (2017), la discrimination et la reproduction de la durée des mots chez les ES (Jones & Yee, 1997).

Les capacités de discrimination et de reproduction relative des durées d'un matériel verbal dans un contexte d'interaction régulière semblent être des capacités temporelles préservées chez les ES mais qui, dans le cadre d'une interaction plus naturelle et donc moins prévisible d'un point de vue temporel, mériteraient d'être renforcées afin d'atteindre un niveau similaire à celui des enfants NE.

b. Perception temporelle moins précise de la régularité de la pulsation

Les résultats de l'expérience n°2 et 3 mettent en évidence chez les ES une moindre précision dans leur perception d'une pulsation dans la parole comme dans la musique.

La mesure de MMN de l'expérience n°2 nous a permis d'analyser la sensibilité des ES à la violation d'une régularité temporelle dans l'échange de parole, autrement dit leur capacité à développer un processus de prédiction temporelle dans l'interaction. La MMN est en effet une réponse évoquée reflétant de manière indirecte la capacité du cortex auditif (associé au cortex frontal) à prédire l'occurrence d'un événement acoustique régulier, que la régularité se manifeste au niveau spectral ou temporel. Ce marqueur électrophysiologique peut être élicité lorsqu'un stimulus acoustique apparaissant de manière régulière est omis, retardé, anticipé ou altéré dans sa forme, et ce, sans qu'une attention sélective soit portée au signal auditif ; la MMN peut par exemple être élicitée pendant le sommeil (pour une revue voir Garrido, Kilner, Stephan, & Friston, 2009). La MMN est ainsi considérée comme le reflet d'une perception inconsciente de la violation d'une régularité acoustique qui est également retrouvée lorsque la violation est produite à partir d'un matériel verbal (Pihko, Leppäsaari, Leppänen, Richardson, & Lyytinen, 1997; Sivonen, Maess, Lattner, & Friederici, 2006). Dans l'expérience n°2, les P-centers des mots du partenaire virtuel étaient distribués à des intervalles réguliers donnant ainsi la pulsation de l'alternance. Dans cette tâche, la stricte isochronie des P-centers permet de considérer la détection de la violation de la régularité de l'alternance, comme la perception d'une déviation temporelle dans la pulsation d'un morceau de musique. L'effet de MMN/MMN-like dans la condition de stimulation auditive va dans le sens des résultats de Torppa et collaborateurs (2014) et de Timm et collaborateurs (2012)

retrouvés chez des enfants et adultes IC respectivement : les ES perçoivent la déviation temporelle de manière moins franche que les enfants NE (Timm et al., 2012; Torppa, Huotilainen, Leminen, Lipsanen, & Tervaniemi, 2014).

Plusieurs facteurs d'ordre perceptif pourraient expliquer la moindre amplitude de la MMN chez les ES. Dans notre tâche, la régularité de l'alternance a été construite à partir d'intervalles temporels plus longs que la fréquence rythmique préférentielle qui est centrée autour de 1.5 Hz (Drake, Jones, & Baruch, 2000). Or, les intervalles réguliers des P-centers du partenaire virtuel sont espacés de 2000 ms soit 0.5 Hz. Etant donné que les enfants possèdent un tempo spontané plus rapide que celui des adultes (McAuley et al., 2006), il est donc possible que cet intervalle temporel ait amoindri la perception de la régularité chez les ES. Cependant, l'amplitude de la MMN retrouvée chez les enfants NE reflète les capacités de perception d'une régularité chez l'enfant entre 6 et 10 ans malgré la fréquence d'apparition lente des stimuli (pour le range perceptif des fréquences rythmiques chez l'adulte voir London, 2004). La perception d'une régularité dans l'alternance n'a peut-être été possible seulement grâce à la participation des enfants à l'interaction verbale. En effet, la dénomination des enfants entre deux P-centers du partenaire virtuel a subdivisé les longs intervalles temporels en deux parties, ce qui a possiblement permis aux enfants de conserver la perception de la régularité en dépit de la fréquence lente de l'alternance imposée par le partenaire virtuel (voir figure 5.2). En outre, les enfants qui ont été les plus précis dans les intervalles irréguliers sont aussi ceux qui ont le mieux perçu l'irrégularité temporelle suggérant un lien entre le placement de la parole de l'enfant et sa perception de l'irrégularité dans l'alternance.

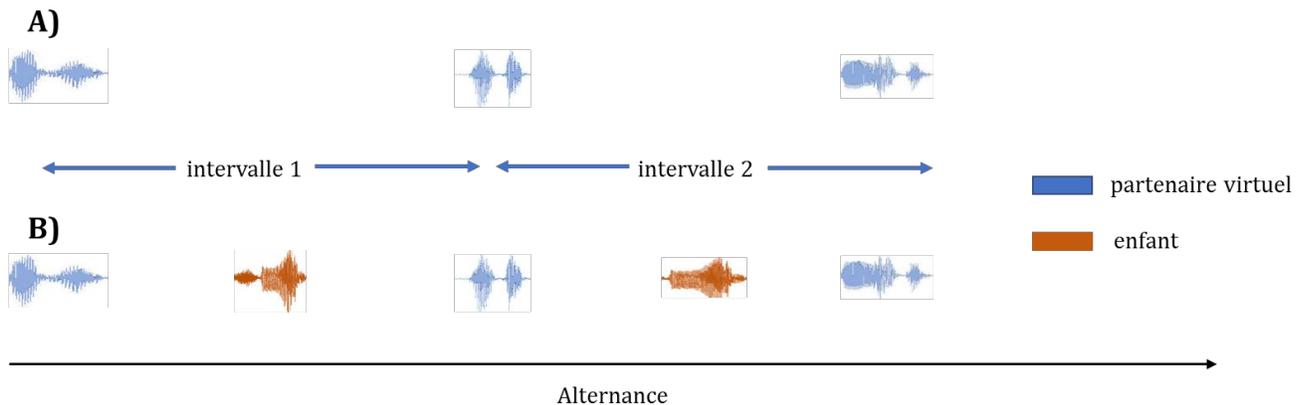


Figure 5.2. Possible facilitation de la perception d'une régularité grâce à l'alternance de parole. Les signaux acoustiques bleus correspondent à la parole du partenaire virtuel (PV). Les signaux acoustiques orange correspondent à la parole de l'enfant. Dans la condition A), les enfants perçoivent la régularité selon la parole du PV seul : les intervalles temporels 1 et 2 entre les mots sont longs et la régularité difficile à maintenir. Dans la condition B) les enfants perçoivent la régularité selon les intervalles temporels entre leur parole et celle du PV : les intervalles temporels entre les mots sont plus courts, la régularité est alors moins difficile à maintenir et la déviation temporelle du PV mieux détectée.

L'hypothèse d'une moindre perception de la régularité chez les ES est renforcée par le déficit de synchronisation des ces enfants dans la tâche de tapping sur métronome. Dans cette tâche, non seulement le stimulus acoustique est d'une composition spectrale plus simple que celle des P-centers, ce qui exige un traitement moins complexe, mais en plus, les intervalles temporels sont de 600 ms qui est une fréquence optimale pour la perception d'une régularité (Repp & Su, 2013). La prise en considération de ces deux paramètres nous permet d'éliminer une difficulté de traitement spectro-temporel dans la réalisation de la tâche ainsi qu' une difficulté d'une perception de la pulsation à cause d'Inter Onset Interval trop longs. Les résultats montrent que les ES ne sont malgré tout pas capables d'être aussi réguliers, donc de prédire la pulsation avec autant de précision que les enfants NE dans la tâche de synchronisation sur métronome (voir Phillips-Silver et al., 2015 pour un résultat similaire chez l'adulte post-lingual implanté cochléaire), alors qu'ils possèdent des capacités de

discrimination des intervalles de durées équivalentes à celles des enfants NE (cf. résultats expérience n°1) et que l'implant cochléaire préserve la transmission des informations temporelles.

Selon la Dynamic Attending Theory (DAT), la perception de la pulsation en musique dépendrait de la synchronisation ou *entrainment* entre des oscillateurs neuronaux et un stimulus extérieur régulier (Large & Jones, 1999). La précision des prédictions temporelles serait basée sur qualité de l'*entrainment*, i.e. l'alignement de phase et de période de l'activité oscillatoire de l'ensemble des neurones du cortex auditif sur le signal acoustique. La MMN étant un indicateur de la sensibilité du système auditif à la régularité et aux prédictions (Bendixen, Sanmiguel, & Schröger, 2012), il semble que ces deux résultats (i.e. moindre amplitude/latence de la MMN et déficit de synchronisation sur métronome) reflètent une altération de la précision avec laquelle les ES perçoivent la régularité d'apparition des stimuli et que le mécanisme déficitaire sous-jacent serait un moindre couplage (phase et/ou période) entre le système auditif et les stimuli réguliers; cette altération de la qualité de l'*entrainment* engendrerait des prédictions temporelles moins précises comparé aux enfants NE dans la parole ou à partir d'un stimulus acoustique simple.

Cependant, la perception de la pulsation est améliorée par la capacité à extraire une structure temporelle de nature hiérarchique des stimuli acoustiques (London, 2004 ; Jones, 2016) et nos résultats montrent que les ES présentent un déficit avéré de cette compétence ce qui pourrait également expliquer la moindre précision avec laquelle ils perçoivent la pulsation. Nous allons discuter de ce point dans la partie suivante.

c. Difficultés de perception d'une structure temporelle de nature hiérarchique

Les résultats des expériences n°1 et n°3 laissent penser que les difficultés de perception temporelle des ES ont pour origine un déficit spécifique du traitement des structures temporelles complexes.

Dans l'expérience n°1, l'absence d'effet de la structure métrique des mots échangés sur la production des ES suggère que contrairement aux enfants NE, les ES, dans la condition de stimulation linguistique, ont des difficultés à percevoir des patterns accentuels dans les mots et/ou à utiliser la régularité de cette structure afin de conserver une consistance dans

l'alternance de parole. Ce résultat va dans le sens des résultats de l'étude de Cason et collaborateurs (2015). Dans cette dernière étude portant sur la capacité de répétition de phrases des ES, même si ces derniers ont amélioré leurs productions verbales après avoir entendu et répété une amorce rythmique (comparé à la ligne de base sans amorçage), la similitude parfaite entre la structure métrique de l'amorce et celle de la phrase à répéter (condition *match*) n'a pas amélioré les performances de ces enfants de manière significative comparé à la condition dans laquelle la structure métrique de l'amorce rythmique était différente de celle de la phrase (condition *mismatch*) (Cason et al., 2015). Il est possible que l'absence d'effet de la structure métrique dans l'expérience n°1 comme dans l'étude de Cason et collaborateurs (2015) provienne des difficultés des ES à percevoir les accents contenus dans les mots. Ces difficultés de perception pourraient être dues à la limitation de la transmission des variations de la fréquence fondamentale via l'implant cochléaire (Torppa et al., 2014). Cependant, l'étude de Segal et collaborateurs (2016), a montré que l'implant fournissait suffisamment d'informations pour permettre la discrimination de mots avec des patterns accentuels différents (Segal, Houston, & Kishon-Rabin, 2016). En outre, les résultats de l'expérience n°3 montrent que malgré l'absence de complexité acoustique du stimulus, les ES présentent des difficultés de synchronisation sur des rythmes complexes qui demandent, comme la perception d'une régularité des patterns accentuels dans les mots, l'extraction plus abstraite, de plusieurs niveaux de régularité hiérarchisés. La construction de ces différents niveaux hiérarchiques est déterminée dans la musique comme dans la parole, par la perception de patterns accentuels autrement dit par la capacité à grouper des événements forts/faibles à différentes échelles temporelles. Le déficit des enfants implantés cochléaires à percevoir des séquences d'évènements (i.e. des patterns) qui se répètent de manière régulière a déjà été évoqué par Conway (2011). L'hypothèse de ces auteurs, nommée *Auditory Scaffolding Hypothesis*, propose que la capacité à apprendre des séquences d'évènements basée sur leur fréquence d'apparition dans le flux acoustique – autrement dit l'apprentissage statistique - est altérée par l'absence de stimulations auditives précoces et pourrait être étendue à toutes les autres modalités sensorielles. L'audition serait en effet par nature, la modalité sensorielle privilégiée des relations entre les éléments qui se déroulent de manière séquentielle, en d'autres mots, l'audition serait le "sens" du traitement temporel. Même si deux nouvelles études (Hall et al., 2017; von Koss Torkildsen et al., 2018) n'ont pas

répliqué les résultats de Conway et collaborateurs, l'altération de cette capacité de structuration de l'information temporelle indispensable au traitement du langage (Romberg & Saffran, 2010) semble, dans nos expériences (1 et 3), être mise en évidence chez ces enfants (dans la modalité auditive), que le stimulus soit un stimulus simple d'un point de vue acoustique telles que les structures rythmiques, ou plus complexe tel que le langage. En outre, le déficit de synchronisation sur la musique va également dans le sens d'une altération de la capacité à structurer l'information auditive. Le stimulus musical est en effet, comme le langage, un stimulus qui contient plusieurs niveaux de régularités (voir chapitre 1) et nécessite, pour se synchroniser, des capacités de traitement temporel plus complexes (i.e. de nature hiérarchique) que celles nécessaires pour réaliser une tâche de tapping sensori-moteur simple. Or, la tâche qui requiert le plus haut niveau de structuration de l'information auditive mais qui demeure peu exigeante d'un point de vue du traitement du signal - i.e. tapping sur des rythmes complexes - est aussi la tâche qui permet le mieux de prédire les niveaux de langage des enfants sourds (expérience n°3).

Ce dernier résultat renforce l'importance de la prise en compte de la nature des opérations de traitements dans le processus de perception auditive ainsi que de la hiérarchie des différentes étapes de traitement. Selon la théorie de la Gestalt notre perception première d'un objet, qu'il soit visuel ou sonore est une perception globale. Autrement dit, la perception de la forme, de la structure d'un objet s'impose à chaque individu selon les lois universelles de la perception bien avant l'analyse de chaque élément qui le compose (Merleau-Ponty, 1945). Cependant, cette forme globale serait plus que le résultat de la somme de ses parties. En d'autres termes, il ne suffirait pas d'assembler les différentes parties d'un objet pour parvenir à s'en faire une représentation signifiante, il faudrait également disposer de la capacité à percevoir le sens global qui les relie. En transposant cette théorie dans le domaine de la perception auditive et la compréhension du langage, on comprend qu'il ne suffirait pas de bien percevoir les paramètres acoustiques de la parole pour parvenir à une représentation pertinente du message verbal. Il faudrait également percevoir la relation dynamique existant entre les éléments de parole dans le contexte d'énonciation donné. Une théorie plus récente (2008), nommée théorie inverse de l'apprentissage perceptuel ou *Reverse Hierarchy Theory* et élaborée suite à des études réalisées dans le domaine visuel, tente d'expliquer la hiérarchie

et la dynamique des traitements par lesquels passe l'information sensorielle avant d'arriver à la perception d'un objet signifiant. Ainsi selon cette théorie, notre système perceptif irait "à l'essentiel". En d'autres termes, notre système perceptif fournirait immédiatement à notre conscience une image des entités de haut niveau qui sont pertinentes dans l'environnement donné (ou pour la réalisation d'une tâche) sans passer par un traitement conscient des différents éléments qui composent ces entités. Notre système tiendrait compte du contexte pour projeter à notre conscience une image pertinente des stimuli perçus. De cette manière, lorsque le signal acoustique est ambigu à cause d'un masquage partiel, il est alors beaucoup plus facile de procéder à une analyse détaillée des éléments masqués puisque ceux-ci sont déjà « contenus » dans l'image globale du message qui a été préalablement projetée. Selon cette théorie, la perception des entités - i.e. niveau de traitement global - serait générée par les cortex associatifs alors que le traitement des unités de bas niveau - i.e. niveau local - serait réalisé dans les cortex sensoriels primaires et le phénomène de perception résulterait d'un va et vient entre les niveau globaux et locaux (voir Ahissar, Nahum, Nelken, & Hochstein, 2009 pour le visuel et Shamma, 2008 pour l'audition). Comme le montrent Farbood et collaborateurs, (2015) dans une étude en fMRI, en langage comme en musique, les structures temporelles les plus élevées dans la hiérarchie - i.e. les plus abstraites - sont traitées par les aires de haut niveau alors que les structures les plus simples sont traitées par le cortex auditif primaire (Farbood, Heeger, Marcus, Hasson, & Lerner, 2015). Gutschalk et collaborateurs (2008) ont également montré via une tâche de masquage informationnel associée à des mesures électrophysiologiques (MEG), que la détection de cibles tonales, donc la prise de conscience de l'apparition d'un signal spécifique dans des conditions d'écoute difficiles, élicite des réponses évoquées tardives dans le système auditif alors que les autres sons (ainsi que tous les sons cible) engendrent tous des réponses dans le cortex auditif primaire (Gutschalk, Micheyl, & Oxenham, 2008). En d'autres termes, la discrimination de deux flux acoustiques concurrents ne devient consciente et pertinente pour l'auditeur que si ce dernier parvient à réaliser des traitements de plus haut niveau en premier lieu (i.e. rechercher, porter une attention à une information qui se dégage du bruit de fond) pour pouvoir ensuite préciser l'analyse de l'information auditive. Il se produirait le même phénomène lorsqu'il s'agit de traiter la parole lors de conversations à plusieurs locuteurs. Ding & Simon (2012) ont en effet montré que la compréhension du message verbal d'un locuteur dans le cadre de voix

concurrentes est améliorée grâce à l'attention spécifique que l'auditeur porte sur le flux de parole du locuteur cible. Les résultats retrouvés suite à des mesures complémentaires en MEG montrent également que l'attention que porte l'auditeur à la parole du locuteur cible, autrement dit l'attente de l'émergence de cette parole, module la représentation neuronale de son flux de parole. En effet, même dans des conditions de faible rapport signal sur bruit (i.e. flux de parole concurrent très présent), l'auditeur encode l'enveloppe de la parole du locuteur cible avec précision (Ding & Simon, 2012). Golumbic et collaborateurs, 2013 ont également montré en mesurant la cohérence de phase entre l'enveloppe temporelle des signaux de parole de locuteurs concurrents (les deux locuteurs parlent en même temps), que les régions de bas niveaux du cortex auditif des auditeurs (gyrus temporal supérieur) se mettent en phase avec les fluctuations lentes (basses fréquences encodant l'enveloppe temporelle: syllabes, mots, phrases prosodiques) et les fluctuations rapides (haut gamma encodant la structure fine : information spectrales) de la parole des deux locuteurs alors que les régions qui traitent les informations linguistiques de haut-niveau (gyrus frontal inférieur, lobule pariétal inférieur et gyrus temporal inférieur et antérieur), se mettent en phase uniquement avec les fluctuations lentes du signal de parole du locuteur cible suggérant une sélectivité du traitement de l'information le long de la hiérarchie du traitement auditif. Les aires de haut niveau (motrices et pariétales), en suivant les fréquences lentes de l'enveloppe du signal, sélectionneraient pour les aires auditives primaires les informations les plus pertinentes à traiter (Golumbic et al., 2013).

Les résultats de ces études ainsi que les théories de la perception décrites ci-dessus suggèrent que percevoir une structure globale dans un contexte donné et y porter une attention spécifique, sont deux processus perceptifs qui ont un impact sur l'encodage et le traitement des stimuli auditifs à bas niveau.

Or, les résultats obtenus par les ES dans l'expérience n°1 associés à leur déficit en tapping sur les rythmes complexes et en musique suggèrent que ces enfants ne parviennent pas à extraire des patterns d'accents (*chunks*) en langage comme en musique et/ou à percevoir une structure temporelle dans le signal acoustique lorsque celle-ci nécessite la mise en relation des éléments selon un mode hiérarchique. Ces enfants semblent utiliser une stratégie cognitive plus coûteuse qui serait de traiter d'abord les éléments qui composent la

structure de manière discrète plutôt que de passer par une perception globale des structures pertinentes pour la tâche. Cette différence de stratégie entre ES et enfants NE peut être analysée dans le cadre de la notion de codage prédictif de l'information ou *predictive coding*. Cette notion, développée par Karl Friston (Friston, 2003) complète l'hypothèse des théories de la Gestalt et de la *reverse hierarchy* en postulant que la perception n'est pas un processus passif mais bien une construction active de notre cerveau. Selon cette notion, le cerveau est en effet considéré non pas comme une unité qui reçoit de l'information sensorielle puis la traite mais comme un système qui génère activement des a priori (*priors*) sur les informations sensorielles qu'il reçoit et qui vérifie sans cesse si les hypothèses qu'il a générées (i.e. erreurs de prédiction) sont justes ou fausses afin de réajuster son modèle interne de la réalité. Ainsi, grâce à ce type d'opérations, le système perceptif génère des modèles de plus en plus précis de l'environnement, qui s'établissent à partir de plusieurs paramètres du contexte dans lequel les stimuli sont produits ce qui permet de les traiter (ou pré-traiter) avant même qu'ils ne parviennent au système perceptif. Dans le cadre de la perception du langage, on peut considérer que le cerveau, au fur et à mesure de la réception du flux continu de parole, construit des hypothèses à propos de la structure et du contenu du langage grâce à l'information précédemment accumulée. Ce type de processus basé sur des prédictions à différentes échelles temporelles permettrait à l'enfant, en actualisant sans cesse son modèle, de maximiser sa perception en minimisant ses erreurs de prédictions. Dans le cadre de la surdit , le m canisme de pr diction pourrait jouer un r le particuli rement important dans le processus de perception car il permettrait de r duire le co t cognitif engendr  par une strat gie de traitement qui serait majoritairement de nature *bottom-up*. Or, nos r sultats sugg rent que deux types de pr dictions temporelles sont alt r s chez l'enfant sourd : la pr diction de la pulsation et de la m trique. Cependant, si les premi res pr dictions sont plus bas es sur la capacit  d'*entrainement/synchronisation*, les pr dictions bas es sur la perception de patterns m triques semblent plus alt r es, probablement car ce type de perception n cessite des capacit s de pr diction qui impliquent le groupement et l'organisation des  v nements en structure.

Concernant l'alt ration de la perception de la pulsation, il est possible de penser que malgr  une transmission de qualit  des informations temporelles par l'implant cochl aire,

l'absence de stimulation précoce voire in utéro ait engendré une réduction de la quantité de substance blanche le long des voies auditives (Kral & Eggermont, 2007). La substance blanche est en effet composée d'axones (i.e. les fibres nerveuses qui conduisent le signal électrique de neurones en neurones) dont la maturation se termine par le processus de myélinisation. Or, le développement de ce processus par lequel les cellules gliales s'entourent autour des axones et accroît la vitesse de conduction des potentiels d'action, commence avec le début de l'entrée en fonction du système auditif (voir Chapitre 3 Eggermont & Moore, *Human Auditory Development*, 2012). La gaine de myéline garantissant la vitesse de propagation de l'information à travers les neurones, on peut penser que la réduction de sa quantité ne permet pas au phénomène d'entraînement de s'établir avec autant de précision que chez les enfants NE et que la perception de la régularité des ES en musique ou en langage n'est alors pas suffisamment robuste pour parvenir à des niveaux de consistance et de détection d'une irrégularité temporelle équivalents à ceux des enfants NE. Par ailleurs, la déprivation sensorielle et le phénomène de réorganisation corticale ont également pu, pour les mêmes raisons de développement neurophysiologique expliquées ci-dessus, diminuer la mise en place des processus top-down, passant par les voies cortico-fugales (Kral & Eggermont, 2007). Cependant, concernant le déficit de prédiction de la métrique, il est probable que le développement des connexions fronto-temporales (Cope et al., 2017) ait également été altéré chez ces enfants à cause de la déprivation sensorielle. On sait en effet que la perception et la prédiction des structures métriques en musique est dépendante de l'expérience auditive (Trehub & Hannon, 2006). On sait d'autre part que la génération de prédictions de haut niveau, notamment dans le cadre de la perception de la parole, passe par une activation des aires frontales (en particulier le gyrus frontal inférieur) (Sohoglu, Peelle, Carlyon, & Davis, 2012) avec une activation plus importante dans des conditions d'écoute dégradées (Davis & Johnsrude, 2003; Shahin, Bishop, & Miller, 2009). Ce type de prédictions passe également par une activation des aires motrices (aire motrice supplémentaire, cortex prémoteur) qui transfèrent, via la voie dorsale, des informations aux aires temporelles auditives (Park, Ince, Schyns, Thut, & Gross, 2015 ; voir aussi le Dual Stream Prediction Model de Tian, Zarate, & Poeppel, 2016). Par ailleurs, une récente étude en imagerie par résonance magnétique fonctionnelle sur 37 enfants sourds sévères à profonds a montré que la préservation de l'activité des structures frontales en pré-implantation est le meilleur

prédicteur des performances langagières après 6 mois d'implantation (Feng et al., 2018). Ceci suggère que suite à la déprivation sensorielle auditive, la voie dorsale (audio-motrice), reliant les structures frontales aux structures temporales est possiblement moins fonctionnelle chez les ES dont les difficultés d'acquisition du langage perdurent. La figure 5.3 (Kral, Kronenberger, Pisoni, & O'Donoghue, 2016) montre en effet que les capacités de fonctionnement du système auditif, qui reposent en partie sur la communication avec les aires frontales, sont dépendantes de l'expérience auditive reçue et que même si une prise en charge précoce peut rétablir la trajectoire développementale de la personne sourde en la rapprochant de celle d'une personne normo-entendante, le niveau de fonctionnalité atteint, surtout si la surdité est congénitale, n'est jamais identique à celui d'une personne normo-entendante.

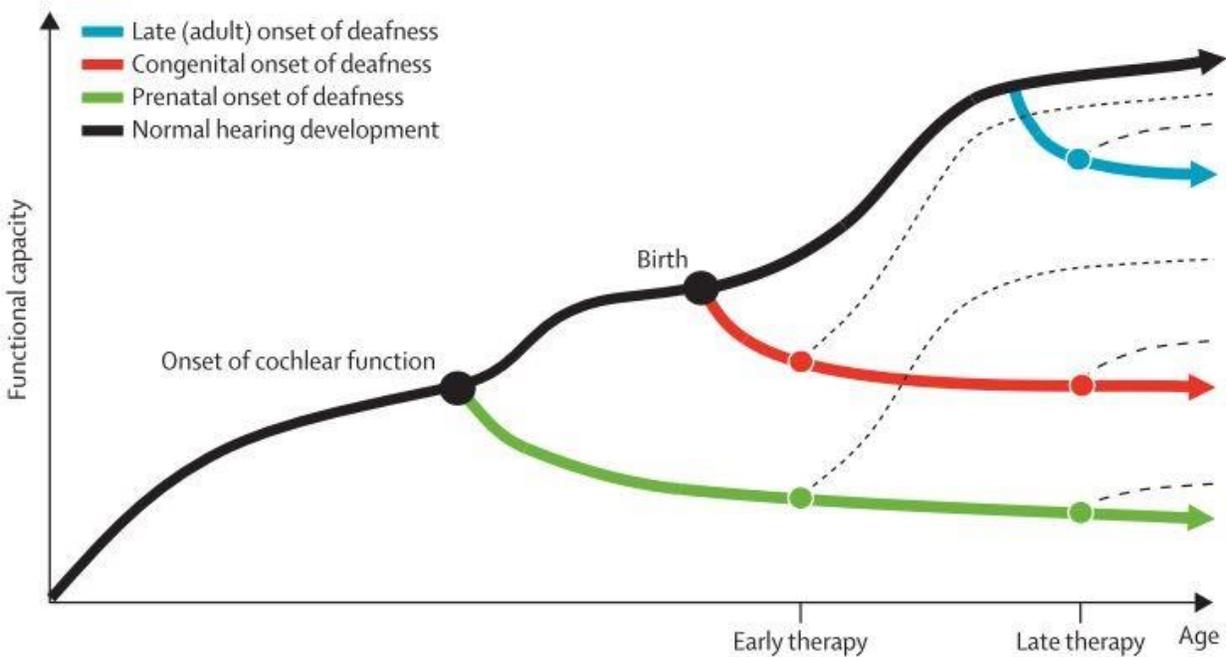


Figure 5.3. Représentation du développement de la capacité fonctionnelle du système auditif en fonction de l'expérience auditive extraite de Kral et collaborateurs (2016). En noir est représenté le développement des capacités du cortex auditif d'une personne normo-entendante : le fonctionnement augmente avec les stimulations reçues. En vert, rouge et bleu, sont représentées les déviations de la trajectoire développementale en fonction de

l'âge de survenue de la surdité. En pointillés, sont représentées les possibles évolutions des capacités fonctionnelles du cortex auditif en fonction de la période de prise en charge.

En conclusion sur les préservations/déficits des capacités de traitement temporel des enfants sourds, il est avéré que le traitement de la durée des intervalles est préservé dans des conditions d'écoute optimales dans lesquelles la durée des stimuli est régulière et demande peu de précision (expérience n°1). Cependant, lorsqu'il s'agit d'utiliser la perception de la durée d'un matériel verbal dans des conditions où celle-ci varie (Freeman et al., 2018), ou bien de régulariser les durées afin d'extraire une pulsation et effectuer une opération cognitive précise (expérience n°2 et n°3) les enfants sourds semblent plus en difficulté. Cette diminution des performances en fonction de la complexité de la tâche suggère que ces enfants possèdent des capacités d'*entrainement* neuronal sur un stimulus régulier mais que le couplage entre le stimulus et l'enveloppe du signal est moins robuste que celui des enfants NE ce qui semble altérer la précision de leurs prédictions temporelles sur la pulsation d'un stimulus simple ou sur un matériel verbal. Par ailleurs, il semble qu'un traitement temporel requérant un plus haut niveau d'abstraction soit plus particulièrement affecté chez les enfants sourds. Le niveau de perception métrique, nécessitant de structurer les éléments de l'input acoustique de manière hiérarchique, et qui, selon les théories de la perception passerait par des processus de type *top-down*, pourrait impacter de manière plus importante leurs capacités générales de prédiction. C'est en effet une difficulté à percevoir les éléments auditifs en structures organisées et à prédire la régularité de ces structures qui semble expliquer le mieux les difficultés de traitement des caractéristiques spectro-temporelles de la parole chez les enfants sourds, en particulier dans le bruit ou dans les conversations à plusieurs interlocuteurs. Nous allons maintenant discuter des effets bénéfiques apportés par 30 minutes de stimulation rythmique sur certains de ces mécanismes temporels mis en jeu dans l'interaction verbale.

2. Quel type d'amélioration une stimulation rythmique de 30 minutes apporte-t-elle aux enfants sourds en termes de traitement temporel de la parole en situation d'interaction ? Un point de vue selon le modèle ADAM.

Quelques études ont déjà montré les effets d'une stimulation rythmique de courte et de longue durée sur la perception et la production du langage à différents niveaux et dans différentes pathologies.

Il a par exemple été montré qu'un entraînement musical rythmique d'une durée de 8 mois améliore les compétences en conscience phonologique des enfants dyslexiques comparé à un entraînement en arts plastiques (Flaunacco et al., 2015). La présentation de courtes amorces rythmiques préalablement à des tâches linguistiques, a également des effets bénéfiques sur le traitement grammatical des enfants souffrant de troubles spécifiques du langage ou de dyslexie (Przybylski et al., 2013) ainsi que sur le traitement phonologique d'enfants implantés cochléaires (IC) ou appareillés de manière conventionnelle (Cason et al., 2015). En outre, l'association d'un amorçage rythmique (versus des bruits environnementaux) à des exercices de langage améliore chez les enfants IC, leurs performances en jugement de grammaticalité, en répétition de non-mots et leurs capacités d'analyse séquentielle de l'information dans des tâches attentionnelles (Bedoin et al., 2017). Les mécanismes ayant produit les effets positifs des stimulations délivrées dans ces différentes études (stimulations de courtes ou de longues durées), ont la plupart du temps été expliqués selon la *Dynamic Attending Theory*, c'est-à-dire par un meilleur couplage des oscillateurs internes sur l'enveloppe du signal acoustique engendrant une possible augmentation de l'attention portée aux stimuli linguistiques dans les tâches.

Contrairement aux autres études, pendant nos expériences, nous avons délivré pendant 30 minutes une stimulation rythmique systématiquement active et c'est la 1ère fois qu'une étude montre les effets de ce type de stimulation sur des performances linguistiques chez des enfants sourds. En outre, nous avons choisi d'analyser les possibles effets du rythme musical sur une tâche linguistique se rapprochant d'une situation de conversation qui est la

situation dans laquelle les enfants sourds sont quotidiennement confrontés et le plus en difficulté.

La conception de la tâche nous permet d'analyser les effets du rythme sur deux compétences essentielles en situation d'interaction verbale que sont les capacités d'anticipation et d'adaptation.

Selon le modèle ADAM, même si ces deux mécanismes sont interdépendants, la mesure de longueur vectorielle, représentant la consistance des enfants dans la tâche, nous renseignerait plus spécifiquement sur le mécanisme d'anticipation alors que la mesure de l'angle, représentant la précision de la production des enfants par rapport à celle du partenaire virtuel, nous renseignerait plus sur le mécanisme d'adaptation temporelle (Mills et al., 2015).

Dans l'expérience n°1, bien que les rythmes entraînés lors de la stimulation n'étaient pas similaires aux rythmes générés dans l'alternance de parole, il semble que la stimulation rythmique, a plutôt amélioré les capacités d'anticipation des enfants sourds : la longueur du vecteur était plus élevée dans la condition de régularité métrique après le rythme. Il est possible de penser que la stimulation rythmique, en entraînant l'enfant à découper les stimuli auditifs en différentes structures, à répéter ces structures et à échanger de courtes structures temporelles avec un partenaire pour en créer une autre sur une plus grande échelle temporelle (cf. détails sur la stimulation), ait favorisé la représentation des accents non comme des éléments séparés et distribués aléatoirement mais plutôt comme des patterns, avec une distribution régulière. Le développement de ce modèle interne de patterns accentuels (faible - fort pour les mots bisyllabiques ou uniquement fort pour les mots monosyllabiques) aurait permis aux enfants sourds d'anticiper les structures métriques des mots du partenaire virtuel et de renforcer leur perception de la régularité entre les P-centers. La régularité étant perçue plus précisément, la consistance de leur production dans l'échange s'en serait trouvée facilitée.

Par ailleurs, il est également possible que la stimulation rythmique ayant été pratiquée de manière active, couplant systématiquement une perception auditive à un mouvement donc mobilisant pendant 30 minutes les connexions entre les structures

motrices et les structures auditives, a favorisé de manière plus importante les capacités de prédictions temporelles basées sur une meilleure perception métrique (Iversen, Repp, & Patel, 2009) comparé à une stimulation auditive sans implication du système moteur (Cason et al., 2015).

En effet, le modèle ADAM décrit deux types d'anticipation que l'on peut transposer au traitement de la parole en situation d'interaction.

Ainsi on retrouverait chez l'auditeur : une anticipation directe, qui procède d'une imitation motrice endogène de la parole du locuteur au fur et mesure de la perception de son flux de parole (débit de parole, gestes oro-moteurs) et une anticipation indirecte, plus prospective, qui procède de la représentation des intentions de l'interlocuteur, i.e. de ses buts dans l'interaction et actions en rapport avec ces intentions. Ces deux types de prédictions nécessitant l'expérience de la contingence entre les commandes motrices efférentes et les informations sensorielles afférentes (Mukherjee et al., 2018), on peut penser que le renforcement du couplage de la voie audio-motrice pendant la stimulation rythmique a facilité les prédictions temporelles durant la tâche et permis à l'enfant d'augmenter la régularité dans le placement de ses P-centers.

L'expérience n°2 nous a permis de montrer que le rythme améliore les capacités d'adaptation des enfants sourds à des variations temporelles de la parole du partenaire virtuel. Comme pour le mécanisme d'anticipation, le mécanisme d'adaptation peut, selon le modèle ADAM, solliciter un ajustement du comportement selon deux processus : le sujet peut réaliser une correction de phase de son mouvement ou de sa parole, qui est une correction inconsciente, et/ou une correction de période, qui est une opération consciente nécessitant un réajustement de l'horloge interne. Même si notre paradigme ne nous permet pas d'orthogonaliser les deux processus, les résultats montrent une réduction de la mesure de l'angle après la stimulation rythmique lors des intervalles les moins prédictibles, autrement dit une meilleure adaptation au partenaire virtuel ainsi qu'un allongement de la durée de la MMN.

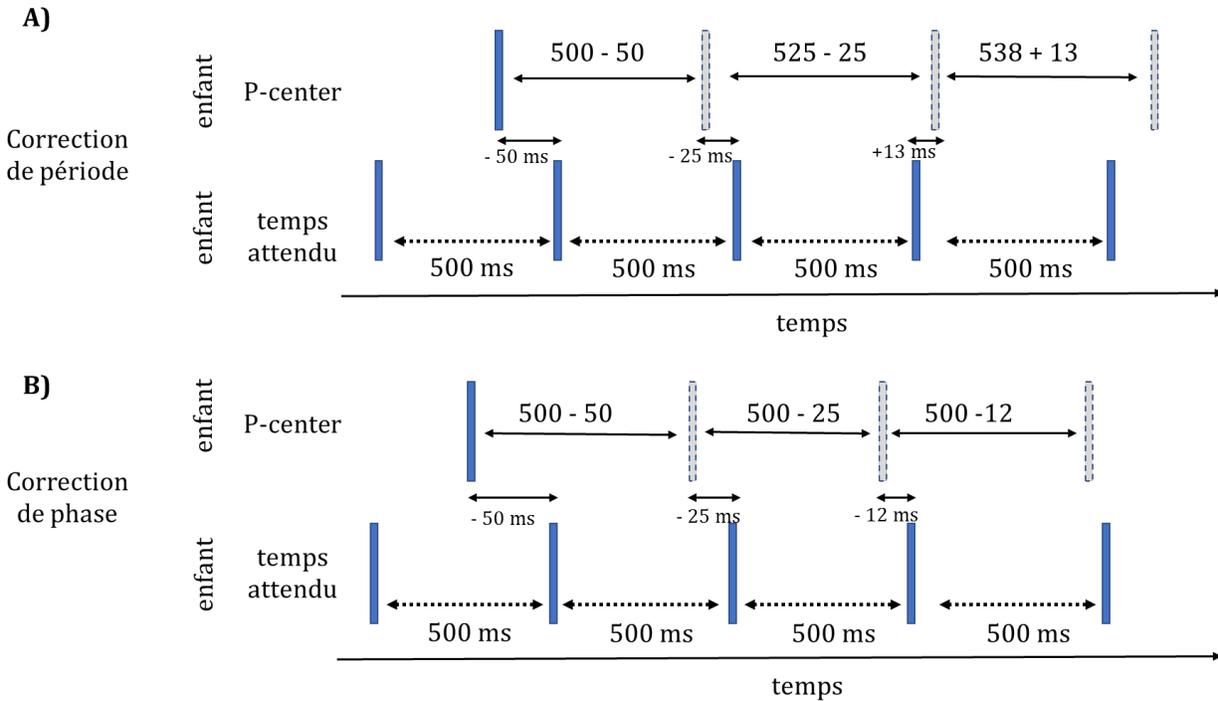


Figure 5.4. Représentation des mécanismes d'adaptation de phase et de période adaptée de van der Steen et Keller (2013).

Panel A) mécanisme de correction de période : il y a un effet cumulatif de la correction de l'erreur qui fait prendre conscience à l'enfant qu'il faut changer de période de référence et ne plus prédire les P-centers sur le tempo qu'il a adopté. Panel B) mécanisme de correction de phase : il n'y a pas d'effet cumulatif de la correction, c'est un ajustement local qui ne nécessite pas de changer de période ou de tempo.

Si l'on considère, selon la notion de *predictive coding*, la MMN comme étant le reflet des prédictions réalisées sur l'input sensoriel (Garrido et al., 2009), on peut penser que l'augmentation de l'amplitude et de la durée de la MMN après la stimulation rythmique signe une amélioration de la mise à jour du modèle interne de la temporalité de l'interaction (Bendixen, Schwartze, & Kotz, 2015). Autrement dit, grâce au rythme, les enfants auraient développé des prédictions plus robustes leur permettant, lors des intervalles irréguliers, de revoir de manière plus évidente les a priori qu'ils avaient formés à propos de l'occurrence temporelle des P-centers du PV. Par ailleurs, l'élicitation d'une MMN-like (P300) peut être considérée comme la perception de la violation d'une régularité à un niveau

conscient (Bekinschtein et al., 2009) c'est-à-dire avec une intention de réaliser une opération cognitive (phase de préparation) ou de prendre une décision motrice (réponse) pertinente dans une tâche (Bolger, Coull, & Schön, 2014). Or, c'est précisément ce type de traitement actif, i.e. prédictions explicites de type *top-down* (Chennu et al., 2013), que la stimulation rythmique semble avoir amélioré et qui est corrélé à l'augmentation de la précision avec laquelle les enfants sont parvenus à s'adapter au PV. On peut aussi penser, à l'instar des résultats retrouvés dans l'étude de Kourtis et collaborateurs (2013), que l'augmentation de la durée de la MMN (i.e. MMN-like) marque la facilitation, par la stimulation rythmique, de la perception de l'échange de paroles comme une action conjointe requérant de la part de l'enfant la planification de ses actions en fonction des contraintes imposées par un partenaire (Kourtis, Sebanz, & Knoblich, 2013).

En résumé, la stimulation rythmique semble avoir favorisé un meilleur couplage stimulus-cerveau, améliorant ainsi la précision de la perception de la régularité de la parole chez les enfants sourds. Ce mécanisme leur aurait permis de mieux s'adapter lors des variations temporelles produites par un partenaire conversationnel. Mais le rythme semble avoir également favorisé des stratégies perceptives différentes de la parole comme le groupement des syllabes en patterns d'accents, leur permettant de percevoir des structures apparaissant régulièrement dans l'échange de parole. En somme, même si l'on ne peut précisément déterminer les mécanismes sous-jacents responsables des effets du rythme dans nos deux expériences, les résultats montrent que le rythme pratiqué de manière active, est une forme de stimulation qui peut moduler le traitement du signal acoustique (Iversen et al., 2009) chez l'enfant sourd en entraînant le développement des prédictions temporelles à un niveau local (expérience n°1) comme global (expérience n°2) ; ce qui est en faveur des théories prônant l'utilisation du rythme dans la remédiation des troubles du langage.

Par ailleurs et d'un point de vue plus clinique, nous avons choisi de délivrer une stimulation de durée moyenne - i.e. 30 minutes - car celle-ci permettait d'exploiter des compétences rythmiques plus diversifiées qu'un entraînement de type amorçage et sur une plus longue durée. Ce type de stimulation est également moins coûteux qu'une étude longitudinale et diminue le risque de perte de patients entre les séances (2 séances ici). En outre, 30 minutes de stimulation est un format qui se rapproche de la durée d'une prise en

charge orthophonique – i.e. 45 minutes en moyenne pour la surdité, ce qui permet d'envisager le transfert de ce type de stimulation dans les conditions habituelles de réhabilitation du langage. La question est de savoir s'il serait préférable de délivrer la stimulation sur une séance entière ou bien de pratiquer des exercices rythmiques en début de séance avant de travailler plus spécifiquement à partir d'un matériel verbal. Selon nous, cela dépend de l'âge et du degré d'engagement de la famille dans la prise en charge. En effet, on peut envisager que la pratique rythmique, en particulier au début de la prise en charge qui est un moment où les parents sont en partie encore sidérés par l'annonce du diagnostic, pourrait être délivrée sur des séances entières afin de rétablir une spontanéité des échanges parent-enfant et de favoriser la fluidité de la communication précoce. On sait en effet que les parents entendants d'enfants sourds sont particulièrement démunis en termes de communication avec leur enfant (Bergeson, Miller, & McCune, 2006) et que la qualité des interactions précoces et en particulier le respect de la rythmicité de ces échanges est un bon prédicteur du développement du langage chez l'enfant normo-entendant (voir chapitre 3) mais également chez l'enfant sourd (Ambrose, Vandam, & Moeller, 2014; Aragon & Yoshinaga-Itano, 2012). Dans le cadre du modèle ADAM, on peut penser que ce type de stimulation musicale non-verbale, pratiqué dans le cadre de prises en charge précoces avec les parents, pourrait faciliter la participation active des parents en contournant la "barrière" linguistique (voire vocale) et restaurer la contingence temporelle des réponses des parents aux sollicitations de l'enfant sourd. Ceci permettrait de sensibiliser l'enfant au paramètre rythmique des échanges et de développer les compétences d'anticipation et d'adaptation nécessaires à la fluidité d'une interaction verbale.

Nous allons maintenant discuter, dans le cadre de la surdité, les hypothèses avancées par les différentes théories qui expliquent les effets de transfert de la musique et du rythme dans la réhabilitation du langage.

3. Est-ce que les théories prônant le rythme/musique comme outil de stimulation pour le langage sont pertinentes dans le cadre de la surdité ?

Les théories de la perception et la notion de codage prédictif, montrent que la faculté à appréhender un stimulus de manière pertinente est dépendante de la capacité à repérer très rapidement des patterns récurrents dans un environnement donné et à réaliser des prédictions sur l'occurrence de ces patterns afin de minimiser l'erreur de prédiction sensorielle ; ce type d'apprentissage revêt un avantage en termes de coût cognitif puisqu'en prédisant de manière de plus en plus précise l'input(acoustique en ce qui nous concerne), le cerveau ne réserve des ressources que pour le traitement des éléments qui demeurent ambigus. Dans le cadre de la surdité, on sait que quel que soit le niveau de surdité, les efforts fournis par les enfants sourds pour désambiguïser le message oral, en particulier dans les situations bruyantes de conversations à plusieurs, est un facteur de fatigue cognitive important (Hornsby, Werfel, Camarata, & Bess, 2014). Il semble donc important de favoriser pour ces enfants des remédiations axées sur la réduction du coût cognitif lié au traitement de l'information acoustique et linguistique en particulier. Or, les opérations cognitives de type *domain-general*, comme l'apprentissage séquentiel et l'attention sélective sont des processus *top-down* nécessaires au traitement de la parole et des théories montrent que la pratique de la musique, et en particulier le rythme, stimulent spécifiquement ces processus en empruntant des réseaux corticaux et en passant par une hiérarchie des traitements similaires à ceux utilisés dans la perception et la production de la parole.

Une théorie, l'hypothèse *OPERA* (Patel, 2011, 2014), formule les conditions nécessaires à un transfert des opérations cognitives réalisées dans le traitement de la musique et du langage, autrement dit à un transfert d'apprentissage entre ces deux activités cognitives. Cette hypothèse explique que les réseaux neuronaux activés lors de la pratique de la musique recouvrent en partie ceux utilisés lors du traitement du langage (Overlap), mais la précision, temporelle et fréquentielle, exigée par la musique est supérieure à celle requise dans le domaine du langage (Precision). La pratique musicale, contrairement au langage emprunte en outre les voies du circuit de la récompense ce qui favorise le désir de renouveler

l'expérience vécue et l'ancrage en mémoire (Emotion) ; ancrage qui est renforcé par la répétition de l'activation des mêmes structures et des éléments acoustiques à l'identique dans le temps (Repetition). De plus, la pratique de la musique oriente spécifiquement l'attention sur les caractéristiques et l'organisation des sons (Attention). Plus précisément concernant le rythme, la *Sound Envelop Hypothesis* (Fujii & Wan, 2014) étend l'hypothèse OPERA et postule que le rythme est un élément musical particulièrement intéressant dans le cadre de la remédiation des troubles du langage car sa perception utilise un phénomène naturel de couplage entre l'activité oscillatoire du cortex auditif et l'enveloppe acoustique du signal musical sur des bandes fréquentielles similaires à celle du langage. De surcroît, sa pratique active comme son écoute passive, permettent de stimuler les mêmes structures (cortico-sous-corticales) que celles dédiées au traitement temporel de la parole (tronc cérébral, cervelet, thalamus, ganglions de la base, aire motrice primaire, aire motrice supplémentaire, cortex préfrontal (cortex préfrontal dorso-latéral, cortex frontal inférieur) et cortex temporal (gyrus temporal supérieur et sillon temporal supérieur). Même si notre but n'est pas d'utiliser des voies de traitement de la parole parallèles, comme dans le cas de patients souffrant de déficits des ganglions de la base, les résultats de l'expérience n°3 nous montrent que les enfants sourds ont besoin de renforcer la qualité du couplage neuronal avec le stimulus de parole afin d'être aussi consistants que les enfants NE et que ce processus semble fonctionner après 30 minutes de stimulation (cf. résultats expérience n°2 MMN). Par ailleurs, la *Precise Auditory Timing Hypothesis* (Tierney & Kraus, 2014) stipule que la qualité de synchronisation sensori-motrice dans les tâches de tapping influence la précision de la perception des indices temporels de la parole tels que la durée des phonèmes, des syllabes ou des mots et renforce l'encodage de leur caractéristiques spectro-temporelles dès les structures sous-corticales comme le colliculus inférieur. Bien que cette hypothèse ait été proposée pour des patients souffrant d'un déficit phonologique se manifestant par un trouble de la lecture, on peut penser qu'un encodage plus précis des sons de la parole à bas niveau permettrait aux enfants sourds de par exemple mieux discriminer les phonèmes dans un environnement bruyant ou de percevoir la différence entre un allongement final marquant une frontière de syntagme accentuel versus une frontière de syntagme intonatif. Ces enfants pourraient ainsi utiliser les indices prosodiques à des fins de compréhension du discours de manière plus efficiente (Holt et al., 2016, 2017). L'hypothèse *Action Simulation*

for Auditory Precision (Patel & Iversen, 2014), étend encore les autres hypothèses et postule que la perception de la pulsation requiert l'utilisation des propriétés de planification temporelle du système moteur ; les prédictions générées dans le système auditif lors de l'écoute du rythme et la modulation des sons traités par le cortex auditif (modulation de la bande Beta) lors des pulsations seraient, selon cette hypothèse, la conséquence de la simulation d'un mouvement périodique, (i.e. un modèle temporel), par le système moteur et de son transfert au système auditif. Dans le cadre de l'interaction, on sait que la génération de modèles internes temporels d'origine motrice est un mécanisme essentiel pour le maintien de la fluidité des échanges (Mandel et al., 2016) et ce mécanisme semble avoir été à l'œuvre dans nos expériences et plus particulièrement dans l'expérience n°2. La probable simulation du débit de parole du PV par le système moteur de l'enfant pourrait expliquer, selon cette hypothèse, la plus grande précision avec laquelle les enfants sourds ont prédit la régularité du PV au point qu'une violation de cette régularité devienne consciente (MMN-like). Enfin, la *Shared Sentence Integration Resource Hypothesis* (Patel, 2003) complète d'une certaine manière l'explication des mécanismes ayant conduit à une amélioration des performances des enfants sourds en particulier dans l'expérience n°1. Même si contrairement à la SEP, cette dernière hypothèse ne fait pas un focus sur le rythme, elle considère que l'intégration des éléments successifs en musique comme en langage sont dépendants d'un mécanisme général, réalisé à partir de ressources neuronales partagées telles que le cortex frontal inférieur et les régions antérieures du cortex temporal supérieur (Koelsch, Gunter, Wittfoth, & Sammler, 2005). Nos résultats montrent que les enfants CI présentent un déficit avéré dans la capacité à intégrer des relations temporelles entre des éléments auditifs qui se succèdent dans le temps même si ces éléments sont simples d'un point de vue acoustique, mais aussi que ces enfants ont développé, grâce au rythme, un encodage plus robuste de la succession des accents comme des structures (patterns) temporelles. Ces deux résultats vont dans le sens, comme l'explique la SSIRH, de ressources partagées lors de ce type de traitement et suggèrent que la pratique rythmique aurait permis aux enfants sourds de traiter les stimuli linguistiques selon une stratégie d'encodage différente, i.e. selon un groupement des unités en musique comme en langage. Par ailleurs, comme le mentionne l'hypothèse OPERA, la répétition, induite par les jeux rythmiques, a probablement sollicité de manière accrue les réseaux spécifiques dévolus au traitement du

langage ce qui peut expliquer en partie, même si elle fut de courte durée, les effets de la stimulation rythmique dans nos tâches. En outre, comme évoqué dans cette même hypothèse, les facteurs émotionnels sont importants pour qu'un transfert d'apprentissage puisse avoir lieu. L'écoute et la pratique du rythme empruntant des circuits cortico (préfrontal) - sous-corticaux (ganglions de la base et tronc cérébral) impliqués dans le circuit de la récompense, on peut penser que la stimulation rythmique, pratiquée de surcroît en interaction avec l'orthophoniste (jeux de partage et d'échanges de rythmes), a favorisé l'engagement et donc l'amélioration des performances des enfants sourds dans la tâche de langage.

Pour résumer, il est possible d'appréhender l'hypothèse OPERA comme la théorie qui résume les conditions qui vont permettre les possibles effets des autres hypothèses.

Les trois hypothèses SEP-PATH-ASAP renforcent l'intérêt de l'implication du système moteur dans l'entraînement auditif pour les enfants sourds car celui-ci permettrait de précisément synchroniser leur cortex auditif sur différentes échelles temporelles présentes aussi dans la parole. La SEP serait à la base de ce processus, via le couplage entre le cerveau et les deux types de stimuli - i.e. l'enveloppe du rythme musical et celle du rythme de parole - ce qui favoriserait la perception de la pulsation ou de la régularité des syllabes dans la parole soit la perception d'une échelle temporelle autour de 200 ms. Selon la PATH, une meilleure synchronisation sur la parole à une échelle temporelle comparable à la pulsation en musique engendrerait un encodage plus robuste des phonèmes dans le colliculus inférieur donc une amélioration de la perception de la parole à une échelle temporelle inférieure à 40 ms. Si le rythme peut avoir un effet à cette échelle temporelle, on peut penser qu'une meilleure synchronisation sur la pulsation de la parole pourrait aussi permettre un meilleur encodage des indices morphosyntaxiques que l'on sait très mal perçus par les enfants IC. La SSIRH se démarque de ces trois hypothèses car elle met l'accent sur un système cérébral commun pour l'intégration de structures dans la musique comme dans la parole au niveau syntaxique. La musique aurait ici un effet sur le traitement du langage à une plus longue échelle temporelle, supérieure à 500 ms ce qui est un point particulièrement important compte-tenu des difficultés syntaxiques des enfants sourds et de leur déficit avéré à dégager une structure temporelle de nature hiérarchique nécessaire au traitement syntaxique.

Par ailleurs, les études appuyant l'hypothèse SSIRH montrent une implication des structures frontales dans ce type de traitement, structures également impliquées dans la planification des actions et les fonctions exécutives telles que la mémoire de travail, et l'on sait que la préservation spécifique de ces structures, observable en pré-implantation, est un bon prédicteur du développement linguistique des enfants IC (Feng et al., 2018).

Le point qu'il manquerait à ces cinq hypothèses dans le cadre de la prise en charge de la surdité serait d'expliquer en quoi la musique et plus particulièrement le rythme permettrait de prédire des structures mais sur de longues échelles temporelles telle que les échelles temporelles prosodiques comme le syntagme intonatif ou encore à des échelles supérieures à 1 seconde. Dans le cadre de l'interaction, les échelles temporelles à prendre en compte sont en effet d'une durée parfois supérieure à 2 secondes et l'enfant est amené à faire des prédictions de l'événement de parole à venir non pas seulement en fonction du contexte immédiat mais aussi à partir d'informations acoustiques et linguistiques perçues bien plus en amont dans l'énoncé. Ces prédictions globales de type Bayésiennes, ou capacités à extraire des régularités statistiques dans le signal sur de longues distances, auraient, elles aussi, des conséquences importantes sur la modulation du traitement de l'information phonémique (Baese-Berk et al., 2014) et serait un atout pour la modulation de l'attention dans un temps plus long que celui de la phrase.

L'acquisition d'une parfaite perception de la régularité ne serait peut-être pas l'élément fondamental pour la perception auditive chez l'enfant sourd car le système auditif semble de toute façon enclin à régulariser les stimuli, mais les capacités de prédictions et de flexibilité temporelles semblent par contre essentielles dans le déroulement du processus de perception (Rimmele et al., 2018) et plus particulièrement dans les situations de conversation. Ainsi, même si les prédictions sont un terme "parapluie" selon Patel et Morgan (2016) qui recouvre différents types de processus (Patel & Morgan, 2016), les inférences de haut-niveau sur de longues échelles temporelles ne sont pas à considérer comme une aide supplémentaire pour comprendre le langage mais sont à appréhender comme le but essentiel de la compréhension du langage en conversation (Kuperberg & Jaeger, 2016).

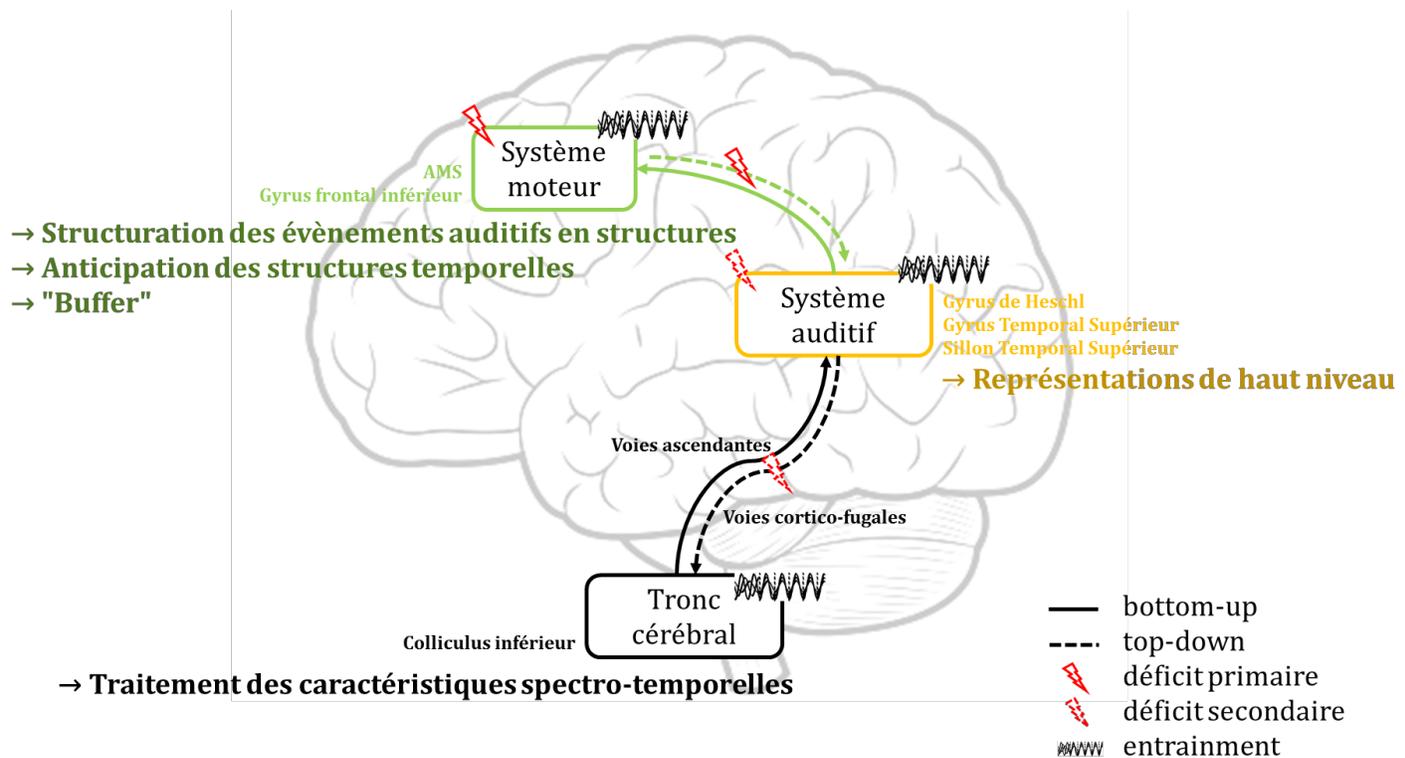


Figure 5.5. Synthèse. Possibles déficits primaires (marqués par un éclair rouge en traits pleins) et secondaires (marqués par un éclair rouge en pointillés) engendrés par la déficience auditive d'un point de vue neuro-cognitif. L'*entraînement* (oscillations en noir) sur l'enveloppe temporelle de la parole interviendrait au niveau de plusieurs structures corticales et sous-corticales. Ce phénomène permettrait la sélection et l'orientation de l'attention sur les entrées sensorielles pour un traitement optimal ainsi que la communication entre les systèmes moteurs (en vert) et auditifs (en jaune) pour le développement de prédictions à plusieurs échelles temporelles : syllabes-mots-phrases-structures prosodiques. La déprivation sensorielle amoindrirait la qualité du couplage entre l'enveloppe temporelle du signal et les oscillations endogènes des différentes structures. Cette moindre synchronisation créerait un déficit de la perception de la pulsation mais surtout altérerait la communication fonctionnelle bidirectionnelle entre les structures audio-motrices (voie dorsale en vert), communication indispensable au développement de représentations abstraites de nature hiérarchique de l'information acoustique. Dans la parole, déjà au niveau du mot, les prédictions sont de nature métrique et nécessitent la perception d'une organisation des patterns accentuels. Au plus les traitements impliquent des éléments hauts dans la hiérarchie

linguistique (phrase, constituants prosodiques, discours), au plus la construction des structures abstraites se complexifie et implique l'activation des structures frontales (en vert). Suite à la déprivation sensorielle, l'insuffisance d'activation des structures frontales pour les entrées auditives pourrait expliquer les déficits de flexibilité et de mémoire de travail auditive des enfants sourds. Ne pouvant suffisamment structurer les informations acoustiques en patterns selon différents niveaux de hiérarchie, les enfants sourds, malgré un bon traitement temporel de la parole via la prothèse ou l'implant cochléaire, ne pourraient conserver en "mémoire-tampon" (idée du gyrus frontal inférieur comme structure "*buffer*" selon Fitch & Martins, 2014) des informations linguistiques organisées, et opérer des prédictions basées sur des régularités statistiques à long terme qui intègreraient de multiples informations contextuelles. La stimulation rythmique, en agissant à la fois sur la qualité de l'*entraînement* mais aussi sur la communication entre les structures corticales qui génèrent l'organisation de l'information auditive, permettrait aux enfants sourds d'encoder les stimuli selon une stratégie différente qui améliorerait la mise en mémoire-tampon (*buffer*) de patterns temporels et leurs capacités de prédiction des informations linguistiques à différentes échelles temporelles pouvant aller jusqu'à des niveaux conversationnels.

En conclusion, même si les théories générales sur la perception nous aident à comprendre pourquoi il ne suffit pas d'améliorer les stratégies de codage des processeurs des IC pour que les enfants sourds parviennent à percevoir et comprendre la parole dans toutes les situations, beaucoup reste à faire autant d'un point de vue théorique qu'expérimental pour répondre à la question du type de stimulation le plus adapté à cette problématique cognitive. Cependant, même si le ou les mécanismes à l'origine des effets obtenus après seulement 30 minutes de stimulation ne sont pas encore élucidés, nos résultats laissent penser que la pratique rythmique est un fort pourvoyeur de plasticité cérébrale chez les enfants sourds qui a des effets de transfert sur le traitement de la parole et pourrait être utilisée de manière plus systématique dans la réhabilitation du langage.

4. Limites et Perspectives

Les objectifs de cette thèse, qui étaient de caractériser les déficits de traitement temporel des enfants sourds et de démontrer l'efficacité d'une intervention de type

rythmique sur leur traitement de la parole en situation d'interaction verbale sont atteints. Il reste maintenant à comprendre quels sont les mécanismes neurocognitifs sous-jacents qui ont permis l'amélioration des compétences des enfants sourds après la stimulation rythmique.

Il faudrait en effet comprendre dans quelle mesure le déficit de perception des structures hiérarchiques est restreint à la modalité auditive ou bien si c'est un déficit multimodal en étudiant par exemple de compétences rythmiques sur des séquences visuelles hiérarchiques.

Il faudrait en outre comprendre le degré de déficit de synchronisation du système auditif de ces enfants en faisant des analyses *stimulus brain coupling* (EEG) sur de la parole et de la musique et mesurer les effets de la pratique musicale sur ce couplage.

Une autre perspective intéressante et prometteuse serait de tester les compétences de prédictions des enfants sourds et les effets de la musique sur ces prédictions dans une tâche de perception uniquement. En effet, nos deux études en production associant EEG et production de parole chez l'enfant n'ont pas toujours facilité le traitement des données et favorisé une situation d'interaction "écologique". C'est pour cela que nous avons mis en place et déjà testé sur une population d'enfants normo-entendants, un nouveau paradigme basé sur la perception de dialogues filmés entre deux personnages au cours duquel nous pouvons analyser les capacités de prédictions des enfants grâce à des mesures de mouvements oculaires ainsi que la qualité de leur compréhension verbale lors des anticipations grâce à des mesures comportementales. Nous avons également testé les effets d'une stimulation rythmique versus tonale mais ces enfants étant de très bons communicateurs, nous avons retrouvé une tendance, mais pas d'effet significatif de la stimulation rythmique. Les enfants sourds étant beaucoup plus en difficulté dans ce type de situation, nous devrions retrouver un effet de la stimulation musicale versus une stimulation orthophonique sur ces capacités nécessaires à la compréhension de la parole en interaction. De plus, nous ajouterons des mesures électroencéphalographiques afin de mesurer leur perception auditive lors des anticipations.

La stimulation rythmique que nous avons proposée montrant des effets au bout de 30 minutes, il serait également intéressant de réaliser une étude longitudinale et pratiquer une évaluation à plus long terme de la stimulation (*follow-up*) afin d'analyser la durée de l'effet de la musique sur la perception et la production de la parole.

Par ailleurs, il faut souligner que la problématique du recrutement des enfants ne nous permet pas d'obtenir un nombre suffisant de données surtout étant donné l'hétérogénéité des profils de ces enfants. C'est pour cette raison qu'il faut entreprendre d'élargir ces recherches à des études multicentriques.

Il semble également important de continuer à analyser les capacités de ces enfants à structurer l'information auditive à haut niveau, il est également important d'augmenter la quantité et la qualité des paradigmes expérimentaux de nature interactionnelle. En effet, la structuration d'une interaction est le socle à partir duquel le langage se développe et le niveau de langage acquis par l'enfant comme le rattrapage éventuel en compétences langagières est dépendant de la qualité de ces interactions. En outre, l'interaction est une situation particulière qui génère des processus de traitement spécifiques, différents de la perception et de la production de parole en situation isolée qui nécessitent d'être pris en compte pour s'approcher le plus précisément possible des conditions d'utilisation du langage des enfants sourds. De plus, les habiletés conversationnelles étant le lieu de l'utilisation sociale du langage, il est indispensable d'explorer plus en détail ces habiletés et d'affiner les techniques de rééducation dans une perspective d'amélioration du développement social de ces enfants.

Références

- Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine.
- Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation Improves Language Comprehension. *Psychological Science*, *21*(12), 1903–1909. <https://doi.org/10.1177/0956797610389192>
- Adank, P., Rueschemeyer, S.-A., & Bekkering, H. (2013). The role of accent imitation in sensorimotor integration during processing of intelligible speech. *Frontiers in Human Neuroscience*, *7*(October), 1–13. <http://doi.org/10.3389/fnhum.2013.00634>
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1515), 285–99. <http://doi.org/10.1098/rstb.2008.0253>
- Ambrose, S. E., Vandam, M., & Moeller, M. P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and Hearing*, *35*(2), 139–147. <http://doi.org/10.1097/AUD.0b013e3182a76768>
- Aragon, M., & Yoshinaga-Itano, C. (2012). Using language ENvironment analysis to improve outcomes for children who are deaf or hard of hearing. *Seminars in Speech and Language*, *33*(4), 340–353. <http://doi.org/10.1055/s-0032-1326918>
- Auer, P. L., Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). Rhythm and Conversational Turn Taking in English. In *Language in time: The rhythm and tempo of spoken interaction*. New York: Oxford University Press.
- Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. a, Morrill, T. H., & McAuley, J. D. (2014). Long-Term Temporal Tracking of Speech Rate Affects Spoken-Word Recognition. *Psychological Science*, *25*(8), 1546–1553. <http://doi.org/10.1177/0956797614533705>
- Bebko, J. M., Calderon, R., & Treder, R. (2003). The Language Proficiency Profile-2: Assessment of the Global Communication Skills of Deaf Children Across Languages and Modalities of Expression. *Journal of Deaf Studies and Deaf Education*, *8*(4), 438–451. <http://doi.org/10.1093/deafed/eng034>
- Bedoin, N., Besombes, A.-M., Escande, E., Dumont, A., Lalitte, P., & Tillmann, B. (2017).

- Boosting syntax training with temporally regular musical primes in children with cochlear implants. *Annals of Physical and Rehabilitation Medicine*. <http://doi.org/10.1016/J.REHAB.2017.03.004>
- Beer, J., Kronenberger, W. G., Castellanos, I., Colson, B. G., Henning, S. C., & Pisoni, D. B. (2014). Executive functioning skills in preschool-age children with cochlear implants. *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(4), 1521–34. http://doi.org/10.1044/2014_JSLHR-H-13-0054
- Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, *106*(5), 1672–1677. <http://doi.org/10.1073/pnas.0809667106>
- Bendixen, A., Sanmiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*, 120–131. <http://doi.org/10.1016/j.ijpsycho.2011.08.003>
- Bendixen, A., Kotz, S. A. & Schwartz, M. (2015). Temporal dynamics of contingency extraction from tonal and verbal auditory sequences. *Brain and Language*, *148*, 64–73. <https://doi.org/10.1016/j.bandl.2014.11.009>
- Bengtsson, S. L., Ullén, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., ... Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex*, *45*(1), 62–71. <http://doi.org/10.1016/j.cortex.2008.07.002>
- Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001–3027. <http://doi.org/10.1016/j.pragma.2011.05.011>
- Bergeson, T. R., Miller, R. J., & McCune, K. (2006). Mother’s speech to hearing-impaired infants and children with cochlear implants. *Infancy*, *10*(3), 221–240. http://doi.org/10.1207/s15327078in1003_2
- Bergeson, T. R., & Trehub, S. E. (2006). Infants Perception of Rhythmic Patterns. *Music*

Perception, 23(4), 345–360. <http://doi.org/10.1525/mp.2006.23.4.345>

Berninger, G., & Garvey, C. (1981). Questions and the allocation, construction, and timing of turns in child discourse. *Journal of Psycholinguistic Research*, 10(4), 375–402. <http://doi.org/10.1007/BF01067165>

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3–4), 399–410. <http://www.ncbi.nlm.nih.gov/pubmed/17943015>

Bolger, D., Coull, J. T., & Schön, D. (2014). Metrical Rhythm Implicitly Orients Attention in Time as Indexed by Improved Target Detection and Left Inferior Parietal Activation. *J Cogn Neurosci.*, 26(3):593-605. http://doi.org/10.1162/jocn_a_00511

Boons, T., De Raeve, L., Langereis, M., Peeraer, L., Wouters, J., & Van Wieringen, A. (2013). Narrative spoken language skills in severely hearing impaired school-aged children with cochlear implants. *Research in Developmental Disabilities*, 34, 3833–3846. <http://doi.org/10.1016/j.ridd.2013.07.033>

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25. [http://doi.org/10.1016/S0010-0277\(99\)00081-5](http://doi.org/10.1016/S0010-0277(99)00081-5)

Breska, A., & Deouell, L. Y. (2016). When Synchronizing to Rhythms Is Not a Good Thing: Modulations of Preparatory and Post-Target Neural Activity When Shifting Attention Away from On-Beat Times of a Distracting Rhythm. *Journal of Neuroscience*, 36(27), 7154–7166. <http://doi.org/10.1523/JNEUROSCI.4619-15.2016>

Caselli, M. C., Rinaldi, P., Varuzza, C., Giuliani, A., & Burdo, S. (2012). Cochlear Implant in the Second Year of Life: Lexical and Grammatical Outcomes. *Journal of Speech Language and Hearing Research*, 55(2), 382. [http://doi.org/10.1044/1092-4388\(2011/10-0248\)](http://doi.org/10.1044/1092-4388(2011/10-0248))

Casillas, M., & Frank, M. C. (2017). The development of children’s ability to track and predict turn structure in conversation. *Journal of Memory and Language*, 92, 234–253. <http://doi.org/10.1016/j.jml.2016.06.013>

- Casillas, M. (2014). Turn Taking. In D. Matthews (Ed) (Vol. 10), *Pragmatic Development in First Language Acquisition* (pp 53-71). Amsterdam: John Benjamins Publishing Company. <http://doi.org/10.1075/tilar.10>
- Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50. <http://doi.org/10.1016/j.actpsy.2014.12.002>
- Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology*, *29*(1), 102–107. <http://doi.org/10.1037/neu0000115>
- Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–8. <http://doi.org/10.1016/j.neuropsychologia.2012.07.018>
- Castellanos, I., Pisoni, D. B., Kronenberger, W. G., & Beer, J. (2015). *Neurocognitive function in deaf children with cochlear implants*. In M. Marschark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies and language: Policy, practice, and research*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190241414.013.17>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.*, *76*, 893–910. Retrieved from <http://dx.doi.org/10.1037/0022-3514.76.6.893>
- Chen, J. K.-C., Chuang, A. Y. C., McMahon, C., Hsieh, J.-C., Tung, T.-H., & Li, L. P.-H. (2010). Music Training Improves Pitch Perception in Prelingually Deafened Children With Cochlear Implants. *Pediatrics*, *125*(4), 793–800. <http://doi.org/10.1542/peds.2008-3620>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*(12), 2844–2854. <http://doi.org/10.1093/cercor/bhn042>
- Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., ... Bekinschtein, T. A. (2013). Expectation and Attention in Hierarchical Auditory Prediction. *Journal of*

- Neuroscience*, 33(27), 11194–11205. <http://doi.org/10.1523/JNEUROSCI.0114-13.2013>
- Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring Neural Entrainment to Beat and Meter in Infants: Effects of Music Background. *Frontiers in Neuroscience*, 10(May), 1–11. <http://doi.org/10.3389/fnins.2016.00229>
- Condon, W. S., & Sander, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science (New York, N.Y.)*, 183(1951), 99–101. <http://doi.org/10.1126/science.183.4120.99>
- Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, 14(1), 69–82. <http://doi.org/10.1111/j.1467-7687.2010.00960.x>
- Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science*, 18, 275–279. <http://doi.org/10.1111/j.1467-8721.2009.01651.x>
- Cope, T. E., Sohoglu, E., Sedley, W., Patterson, K., Jones, P. S., Wiggins, J., ... Rowe, J. B. (2017). Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nature Communications*, 8(1). <http://doi.org/10.1038/s41467-017-01958-7>
- Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. In *Proceedings of speech prosody 2002* (pp. 121–126). Aix en Provence, France
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2), 139–148. [http://doi.org/10.1016/S0095-4470\(02\)00082-7](http://doi.org/10.1016/S0095-4470(02)00082-7)
- Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, 7(1–2), 28–35. <https://kb.osu.edu/dspace/handle/1811/52976>
- D’Ausilio, A. (2009). Mirror-like mechanisms and music. *TheScientificWorldJournal*, 9, 1415–22. <http://doi.org/10.1100/tsw.2009.160>

- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.*, 23, 3423–3431. <https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003>
- De Ruiter, J., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language*, 82, 515–535. <http://dx.doi.org/10.1353/lan.2006.0130>
- Deocampo, J. A., Smith, G. N. L., Kronenberger, W. G., Pisoni, D. B., & Conway, C. M. (2018). The Role of Statistical Learning in Understanding and Treating Spoken Language Outcomes in Deaf Children With Cochlear Implants. *Language Speech and Hearing Services in Schools*, 49(3S), 723. http://doi.org/10.1044/2018_LSHSS-STLT1-17-0138
- Dettman, S. J., Dowell, R. C., Choo, D., Arnott, W., Abrahams, Y., Davis, A., ... Briggs, R. J. (2016). Long-Term communication outcomes for children receiving cochlear implants younger than 12 months: A multicenter study. *Otology and Neurotology* 37, 82–95. <http://doi.org/10.1097/MAO.0000000000000915>
- Di Cristo, A. (2002). De la métrique et du rythme de la parole ordinaire : l'exemple du français. Bordas, E. (éd.) : Le rythme de la prose. *Revue de sémio-linguistique des textes et discours*, 16, pp.25-43.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), 11854–9. <http://doi.org/10.1073/pnas.1205381109>
- Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The Roots of Turn-Taking in the Neonatal Period. *Infant and Child Development*, 25(3), 240–255. <http://doi.org/10.1002/icd.1976>
- Dowell, R. C., Dettman, S. J., Blamey, P. J., Barker, E. J., & Clark, G. M. (2002). Speech perception in children using cochlear implants: prediction of long-term outcomes. *Cochlear*

Implants International, 3(1), 1–18. <http://doi.org/10.1179/cim.2002.3.1.1>

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77(3), 251–288. [http://doi.org/10.1016/S0010-0277\(00\)00106-2](http://doi.org/10.1016/S0010-0277(00)00106-2)

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PloS One*, 5(8), e12166. <http://doi.org/10.1371/journal.pone.0012166>

Echols, C. H., Crowhurst, M. J., Childers, J. B., Becker, M., & Rader, A. (1997). The Perception of Rhythmic Units in Speech by Infants and Adults. *Journal of Memory and Language*, 36 (2). 202-225. <https://doi.org/10.1006/jmla.1996.2483>

Eggermont, J. J., & Ponton, C. W. (2003). Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: Correlations with changes in structure and speech perception. *Acta Oto-Laryngologica*, 123(2), 249–252. <http://doi.org/10.1080/0036554021000028098>

Eggermont J.J. & Moore, J.K. (2011). Morphological and functional development of the auditory nervous system . In Werner, L., Fay, R. R., & Popper, A. (Eds.), *Human Auditory Development*. (PP 61-107). New York: Springer.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399–402. <http://doi.org/10.1046/j.0953-816x.2001.01874.x>

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience*, 9, 157. <http://doi.org/10.3389/fnins.2015.00157>

Feng, G., Ingvalson, E. M., Grieco-Calub, T. M., Roberts, M. Y., Ryan, M. E., Birmingham, P., ... Wong, P. C. M. (2018). Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients. *Proceedings of the National Academy of Sciences*, (January), 201717603. <http://doi.org/10.1073/pnas.1717603115>

- Figueras, B., Edwards, L., & Langdon, D. (2008). Executive function and language in deaf children. *Journal of Deaf Studies and Deaf Education*, 13(3), 362–77. <http://doi.org/10.1093/deafed/enm067>
- Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7(October), 68. <http://doi.org/10.3389/fnsys.2013.00068>
- Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1316(1), 87–104. <http://doi.org/10.1111/nyas.12406>
- Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music Training Increases Phonological Awareness and Reading Skills in Developmental Dyslexia: A Randomized Control Trial. *Plos One*, 10(9), e0138715. <http://doi.org/10.1371/journal.pone.0138715>
- Foti, D., & Roberts, F. (2016). The neural dynamics of speech perception: Dissociable networks for processing linguistic content and monitoring speaker turn-taking. *Brain and Language*, 157, 63–71. <http://doi.org/10.1016/j.bandl.2016.05.001>
- Fowler, C. A., Smith, M. R., & Tassinary, L. G. (1986). Perception of syllable timing by prebabbling infants. *The Journal of the Acoustical Society of America*, 79(3), 814–825. <http://doi.org/10.1121/1.393472>
- François, C., Grau-Sánchez, J., Duarte, E., & Rodriguez-Fornells, A. (2015). Musical training as an alternative and effective method for neuro-education and neuro-rehabilitation. *Frontiers in Psychology*, 6, 475. <https://doi.org/10.3389/fpsyg.2015.00475>
- Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex (New York, N.Y.: 1991)*, 21(10), 2357–65. <http://doi.org/10.1093/cercor/bhr022>
- Freeman, V., & Pisoni, D. B. (2017). Speech rate, rate-matching, and intelligibility in early-implanted cochlear implant users. *The Journal of the Acoustical Society of America*,

142(2), 1043–1054. <http://doi.org/10.1121/1.4998590>

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325–1352. <http://doi.org/10.1016/J.NEUNET.2003.06.005>

Fu, Q.-J., Galvin, J. J., Wang, X., & Wu, J.-L. (2015). Benefits of Music Training in Mandarin-Speaking Pediatric Cochlear Implant Users. *Journal of Speech Language and Hearing Research*, 58(1), 163. http://doi.org/10.1044/2014_JSLHR-H-14-0127

Fujii, S., & Wan, C. Y. (2014). The Role of Rhythm in Speech and Language Rehabilitation: The SEP Hypothesis. *Frontiers in Human Neuroscience*, 8(October), 1–15. <http://doi.org/10.3389/fnhum.2014.00777>

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of Dancing and Singing to Music in Three- to Four-Months-Old Infants. *PLoS ONE*, 9(5), e97680. <http://doi.org/10.1371/journal.pone.0097680>

Fuller, C. D., Galvin, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of Two Music Training Approaches on Music and Speech Perception in Cochlear Implant Users. *Trends in Hearing*, 22, 2331216518765379. <http://doi.org/10.1177/2331216518765379>

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157. <http://doi.org/10.1016/j.newideapsych.2013.03.005>

Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction Studies*, 13, 103–124. <http://doi.org/10.1075/is.13.1.07fus>

Galantucci, B., & Sebanz, N. (2009). Joint Action: Current Perspectives. *Topics in Cognitive Science*, 1(2), 255–259. <http://doi.org/10.1111/j.1756-8765.2009.01017.x>

Galvin, J. J., Fu, Q.-J., Shannon, R. V., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, 1169, 518–33. <http://doi.org/10.1111/j.1749-6632.2009.04551.x>

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4(June), 1–9. <http://doi.org/10.3389/fpsyg.2013.00340>

- Ganek, H., McConkey Robbins, A., & Niparko, J. K. (2012). Language Outcomes After Cochlear Implantation. *Otolaryngologic Clinics of North America*, 45(1), 173–185. <http://doi.org/10.1016/j.otc.2011.08.024>
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463. <http://doi.org/10.1016/j.clinph.2008.11.029>
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1), 8–11. <http://doi.org/10.1016/j.tics.2003.10.016>
- Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, 1(2), 292–304. <http://doi.org/10.1111/j.1756-8765.2009.01020.x>
- Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6(June), 1–12. <http://doi.org/10.3389/fpsyg.2015.00751>
- Geers, A. E. (1997). Comparing implants with hearing aids in profoundly deaf children. *Otolaryngology - Head and Neck Surgery*, 117(3 I), 150–154. [http://doi.org/10.1016/S0194-5998\(97\)70167-0](http://doi.org/10.1016/S0194-5998(97)70167-0)
- Geers, A. E., Moog, J. S., Biedenstein, J., Brenner, C., & Hayes, H. (2009). Spoken language scores of children using cochlear implants compared to hearing age-mates at school entry. *Journal of Deaf Studies and Deaf Education*, 14(3), 371–385. <http://doi.org/10.1093/deafed/enn046>
- Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language Skills of Children with Early Cochlear Implantation. *Ear & Hearing*, 24, 46–58. <http://doi.org/10.1097/01.AUD.0000051689.57380.1B>
- Geers, A. E., & Sedey, A. L. (2011). Language and Verbal Reasoning Skills in Adolescents With 10 or More Years of Cochlear Implant Experience. *Ear & Hearing*, 32, 39–48. <http://doi.org/10.1097/AUD.0b013e3181fa41dc>
- Gfeller, K. (2016). Music-based training for pediatric CI recipients: A systematic analysis of

- published studies. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 133, S50–S56. <http://doi.org/10.1016/j.anorl.2016.01.010>
- Giles, H., Coupland, N., & Coupland, J. (1991). Accomodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accomodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.
- Ginsburg, G. P., & Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *Journal of Child Language*, 15(02), 221. <http://doi.org/10.1017/S0305000900012344>
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100(13), 8030-8035. <http://www.pnas.org/content/100/13/8030.abstract>
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515–523. <http://doi.org/10.1111/j.1467-9280.2008.02117.x>
- Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of Music Training for Perception of Emotional Speech Prosody in Deaf Children With Cochlear Implants. *Ear and Hearing*, 38(4), 455–464. <http://doi.org/10.1097/AUD.0000000000000402>
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner, & C. Gussenhoven (Eds.), *Papers in laboratory phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter.
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, 6, 55–69. <http://doi.org/10.3389/fpsyg.2015.01167>
- Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural Correlates of Auditory Perceptual Awareness under Informational Masking. *PLoS Biology*, 6(6), e138.

<http://doi.org/10.1371/journal.pbio.0060138>

Hadley, L. V., Novembre, G., Keller, P. E., & Pickering, M. J. (2015). Causal Role of Motor Simulation in Turn-Taking Behavior. *The Journal of Neuroscience*, 35(Xx), 11–15. <http://doi.org/10.1523/JNEUROSCI.1850-15.2015>

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, 86(December), 150–165. <http://doi.org/10.1016/j.neubiorev.2017.12.002>

Hall, M. L., Eigsti, I.-M., Bortfeld, H., & Lillo-Martin, D. (2017). Auditory access, language access, and implicit sequence learning in deaf children. *Developmental Science*, e12575. <http://doi.org/10.1111/desc.12575>

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, 16(1), 48–55. <http://doi.org/10.1111/j.0956-7976.2005.00779.x>

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? *Psychological Science*, 15(6), 409–414. <http://doi.org/10.1111/j.0956-7976.2004.00693.x>

Heldner, A., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. <http://doi.org/10.1016/j.wocn.2010.08.002>

Henning, A., & Striano, T. (2011). Infant and Maternal Sensitivity to Interpersonal Timing. *Child Development*, 82(3), 916–931. <http://doi.org/10.1111/j.1467-8624.2010.01574.x>

Hess, C., Zettler-Greeley, C., Godar, S. P., Ellis-Weismer, S., & Litovsky, R. Y. (2014). The Effect of Differential Listening Experience on the Development of Expressive and Receptive Language in Children With Bilateral Cochlear Implants. *Ear and Hearing*. <http://doi.org/10.1097/AUD.000000000000023>

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Rev. Neurosci.*, 8, 393–402. <http://dx.doi.org/10.1038/nrn2113>

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of

- speech rhythm during joint story building. *Frontiers in Psychology*, 6(June), 1–6. <http://doi.org/10.3389/fpsyg.2015.00797>
- Holt, C. M., Demuth, K., & Yuen, I. (2016). The Use of Prosodic Cues in Sentence Processing by Prelingually Deaf Users of Cochlear Implants. *Ear and Hearing*, 37(4), e256–e262. <http://doi.org/10.1097/AUD.0000000000000253>
- Holt, C. M., Yuen, I., & Demuth, K. (2017). Discourse Strategies and the Production of Prosody by Prelingually Deaf Adolescent Cochlear Implant Users. *Ear and Hearing*, 38(2), e101–e108. <http://doi.org/10.1097/AUD.0000000000000362>
- Hornsby, B. W. Y., Werfel, K., Camarata, S., & Bess, F. H. (2014). Subjective fatigue in children with hearing loss: some preliminary findings. *American Journal of Audiology*, 23(1), 129–34. [http://doi.org/10.1044/1059-0889\(2013/13-0017\)](http://doi.org/10.1044/1059-0889(2013/13-0017))
- Houston, D. M., Beer, J., Bergeson, T. R., Chin, S. B., Pisoni, D. B., & Miyamoto, R. T. (2012). The ear is connected to the brain: some new directions in the study of children with cochlear implants at Indiana University. *Journal of the American Academy of Audiology*, 23(6), 446–63. <http://doi.org/10.3766/jaaa.23.6.7>
- Ibertsson, T., Hansson, K., Maki-Torkko, E., Willstedt-Svensson, U., & Sahlen, B. (2009). Deaf teenagers with cochlear implants in conversation with hearing peers. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, 44(3), 319–37. <http://doi.org/10.1080/13682820802052067>
- Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Annals of the New York Academy of Sciences*, 1169(1), 58–73. <http://doi.org/10.1111/j.1749-6632.2009.04579.x>
- Jackson, P. L., & Decety, J. (2004). Motor cognition: A new paradigm to study self-other interactions. *Current Opinion in Neurobiology*. <http://doi.org/10.1016/j.conb.2004.01.020>
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in*

Child Development, 66(2), 1–132. <http://doi.org/10.2307/3181589>

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, 47(2), 735–744. <http://doi.org/10.1016/j.neuroimage.2009.04.090>

Jiam, N. T., Caldwell, M., Deroche, M. L., Chatterjee, M., & Limb, C. J. (2017). Voice Emotion Perception and Production in Cochlear Implant Users. *Hearing Research*. <http://doi.org/10.1016/j.heares.2017.01.006>

Jones, M. R. (2016). Musical time. In *The Oxford handbook of music psychology* (2nd ed). (pp. 125–141). New York, NY, US: Oxford University Press.

Jones, M. R., Moynihan, H., Mackenzie, N., & Puente, J. (2002). Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays, *13*(4), 313–319.

Jones, M. R., & Yee, W. (1997). Sensitivity to time change: The role of context and skill. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 693–709. <http://doi.org/10.1037/0096-1523.23.3.693>

Jun, S.-A., & Fougeron, C. (2002). *Realizations of accentual phrase in French intonation*. *Probus* (Vol. 14). <https://pdfs.semanticscholar.org/7c75/1788787524ca789c0b53a32ef42763e45e58.pdf>

Karpati, F. J., Giacosa, C., Foster, N. E. V., Penhune, V. B., & Hyde, K. L. (2016). Sensorimotor integration is enhanced in dancers and musicians. *Experimental Brain Research*, 234(3), 893–903. <http://doi.org/10.1007/s00221-015-4524-1>

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, 3(1692), 1–8. <http://doi.org/10.1038/srep01692>

Keitel, A., & Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology*,

6(February), 108. <http://doi.org/10.3389/fpsyg.2015.00108>

Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. von, & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology*, 116(2), 264–277. <http://doi.org/10.1016/j.jecp.2013.06.005>

Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development*, 70(2), 474–485. <http://doi.org/10.1111/1467-8624.00034>

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658). <http://rstb.royalsocietypublishing.org/content/369/1658/20130394.abstract>

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16(1), 55–68. <http://doi.org/10.1017/S030500090001343X>

Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, 52(4), 255–289. <http://doi.org/10.1080/0163853X.2014.955997>

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., ... Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior & Development*, 32, 59–71. <http://doi.org/10.1016/j.infbeh.2008.10.002>

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: learning anticipatory control. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 29(5), 1006–16. <http://doi.org/10.1037/0278-7393.29.5.1006>

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, 17(10), 1565–1577. <http://doi.org/10.1162/089892905774597290>

- Kong, Y.-Y., Cruz, R., Jones, J. A., & Zeng, F.-G. (2004). Music Perception with Temporal Cues in Acoustic and Electric Hearing. *Ear and Hearing, 25*(2), 173–185. <http://doi.org/10.1097/01.AUD.0000120365.97792.2F>
- Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *The Quarterly Journal of Experimental Psychology, 63*(11), 2220–2230. <http://doi.org/10.1080/17470218.2010.497843>
- Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: Social interaction modulates action simulation. *Biology Letters, 6*(6), 758–761. <http://doi.org/10.1098/rsbl.2010.0478>
- Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other people's actions in joint action planning: An EEG study. *Social Neuroscience, 8*(1), 31–42. <http://doi.org/10.1080/17470919.2012.694823>
- Kral, A., & Eggermont, J. J. (2007). What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews, 56*(1), 259–269. <http://doi.org/10.1016/j.brainresrev.2007.07.021>
- Kral, A., Kronenberger, W. G., Pisoni, D. B., & O'Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: a connectome model. *The Lancet Neurology, 15*(6), 610–621. [http://doi.org/10.1016/S1474-4422\(16\)00034-X](http://doi.org/10.1016/S1474-4422(16)00034-X)
- Kral, A., Yusuf, P. A., & Land, R. (2017). Higher-order auditory areas in congenital deafness: Top-down interactions and corticocortical decoupling. *Hearing Research, 343*, 50–63. <http://doi.org/10.1016/j.heares.2016.08.017>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat Rev Neurosci, 11*(8), 599–605. <http://dx.doi.org/10.1038/nrn2882>
- Kronenberger, W. G., Pisoni, D. B., Henning, S. C., & Colson, B. G. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology, 38*(8), 902–914. <http://doi.org/10.1093/jpepsy/jst034>

- Kuhl, P. K. (2012). *Social Mechanisms in Early Language Acquisition: Understanding Integrated Brain Systems Supporting Language*.
<http://doi.org/10.1093/oxfordhb/9780195342161.013.0043>
- Kung, S. J., Tzeng, O. J. L., Hung, D. L., & Wu, D. H. (2011). Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. *Experimental Brain Research*, *210*(2), 269–282. <http://doi.org/10.1007/s00221-011-2630-2>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. <http://doi.org/10.1080/23273798.2015.1102299>
- Large, E. W., & Jones, M. R. (1999a). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159.
- Large, E. W., & Jones, M. R. (1999b). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. <http://doi.org/10.1037/0033-295X.106.1.119>
- Lederberg, A. R., Schick, B., Spencer, P. E., A.R., L., B., S., & P.E., S. (2013). Language and literacy development of deaf and hard-of-hearing children: successes and challenges. *Developmental Psychology*, *49*(1), 15–30. <http://doi.org/10.1037/a0029558>
- Lesicko, A. M. H., & Llano, D. A. (2017). Impact of peripheral hearing loss on top-down auditory processing. *Hearing Research*, *343*, 4–13. <http://doi.org/10.1016/J.HEARES.2016.05.018>
- Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, *20*, 6–14. <http://doi.org/10.1016/j.tics.2015.10.010>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(June), 1–17. <http://doi.org/10.3389/fpsyg.2015.00731>

- Levitan, R., Beňuš, Š., Gravano, A., & Hirschberg, J. (2015). Entrainment and Turn-Taking in Human-Human Dialogue. In *AAAI Spring Symposium on Turn-taking and Coordination in Human- Machine Interaction* (pp. 44–51).
- Lieberman, A. F., & Garvey, C. (1977). Interpersonal pauses in preschoolers' verbal exchanges. In *biennial meeting of the Society for Research in Child Development, New Orleans, LA*.
- Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, 308, 13–26. <http://doi.org/10.1016/j.heares.2013.04.009>
- Loehr, J. D., Large, E. W., & Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1292–1309. <http://doi.org/10.1037/a0023102>
- Lomber, S. G., Meredith, M. A., & Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nature Neuroscience*, 13(11), 1421–1427. <http://doi.org/10.1038/nn.2653>
- Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3, 376. <http://doi.org/10.3389/fpsyg.2012.00376>
- Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, 19(23), 1994–1997. <http://doi.org/10.1016/j.cub.2009.09.064>
- Manson, J. H., Bryant, G. a., Gervais, M. M., & Kline, M. a. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34, 419–426. <http://doi.org/10.1016/j.evolhumbehav.2013.08.001>
- Marcus, S. M. (1981). Acoustic determinants of perceptual centre (P-center) location. *Percept. Psychophys.*, 30, 247–256.

- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*(2), 294–305. <http://doi.org/10.1162/jocn.2010.21413>
- Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language*, *20*(2), 303-312. doi:10.1017/S0305000900008291
- Stephens, G. & Matthews, D. (2014). Turn Taking. In D. Matthews (Ed) (Vol. 10), *Pragmatic Development in First Language Acquisition* (pp 53-71). Amsterdam: John Benjamins Publishing Company. <http://doi.org/10.1075/tilar.10>
- McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367. <http://doi.org/10.1037/0096-3445.135.3.348>
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (2002). A precursor of language acquisition in young infants. *Psycholinguist Crit Concepts Psychol*, *4*, 25. [http://doi.org/10.1016/0010-0277\(88\)90035-2](http://doi.org/10.1016/0010-0277(88)90035-2)
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Éditions Gallimard.
- Mills, P. F., Schultz, B. G., van der Steen, M. C. (Marieke), & Keller, P. E., Van Der Steen, M., Schultz, B. G., & Keller, P. E. (2015). Individual Differences in Temporal Anticipation and Adaptation During Sensorimotor Synchronization. *Timing & Time Perception*, *3*(1–2), 13–31. <http://doi.org/10.1163/22134468-03002040>
- Monk, C., Fifer, W. P., Myers, M. M., Sloan, R. P., Trien, L., & Hurtado, A. (2000). Maternal stress responses and anxiety during pregnancy: Effects on fetal heart rate. *Developmental Psychobiology*, *36*(1), 67–77. [http://doi.org/10.1002/\(SICI\)1098-2302\(200001\)36:1<67::AID-DEV7>3.0.CO;2-C](http://doi.org/10.1002/(SICI)1098-2302(200001)36:1<67::AID-DEV7>3.0.CO;2-C)
- Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, *5*, 5255. <http://doi.org/10.1038/ncomms6255>

- Morton, J., Marcus, S. M., & Frankish, C. R. (1976). Perceptual Centers (P-centers). *Psychological Review*, 83(5), 405–408. <http://doi.org/10.1037/0033-295X.83.5.405>
- Most, T., Shina-August, E., & Meilijson, S. (2010). Pragmatic abilities of children with hearing loss using cochlear implants or hearing AIDS compared to hearing children. *Journal of Deaf Studies and Deaf Education*, 15(4), 422–37. <http://doi.org/10.1093/deafed/enq032>
- Mukherjee S, Badino L, Hilt PM, et al. The neural oscillatory markers of phonetic convergence during verbal interaction. *Hum Brain Mapp*. 2018;1 - 15. <https://doi.org/10.1002/hbm.24364>
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Reserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, 2(2), 164–173. <http://doi.org/10.1111/1467-7687.00065>
- Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, 131(2), 1307–14. <http://doi.org/10.1121/1.3672697>
- Nakata, T., Trehub, S. E., Mitani, C., & Kanda, Y. (2006). Pitch and Timing in the Songs of Deaf Children With Cochlear Implants, 24(2), 147–154.
- Natale, M. (1975). Convergence of Mean Vocal Intensity in Dyadic Communication as a Function of Social Desirability. *Journal of Personality and Social*, 1(5), 790–804. <http://doi.org/10.1037/0022-3514.32.5.790>
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756–766. <http://doi.org/10.1037/0096-1523.24.3.756>
- Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41(1), 233–243. [http://doi.org/10.1016/S0167-6393\(02\)00106-1](http://doi.org/10.1016/S0167-6393(02)00106-1)

- Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., ... Team, for the Cd. I. (2010). Spoken Language Development in Children Following Cochlear Implantation. *JAMA*, *303*(15), 1498. <http://doi.org/10.1001/jama.2010.451>
- Novembre, G., & Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Frontiers in Human Neuroscience*, *8*(August), 603. <http://doi.org/10.3389/fnhum.2014.00603>
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. *Social Cognitive and Affective Neuroscience*, *9*(8), 1062–1068. <http://doi.org/10.1093/scan/nst086>
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, *31*(28), 10234–40. <http://doi.org/10.1523/JNEUROSCI.0411-11.2011>
- Knoblich, G., & Sebanz, N. (2006). The Social Nature of Perception and Action. *Current Directions in Psychological Science*, *15*(3), 99–104. <https://doi.org/10.1111/j.0963-7214.2006.00415.x>
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–61. <http://doi.org/10.1097/AUD.0b013e3181b412e9>
- Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia*, *49*(12), 3338–45. <http://doi.org/10.1016/j.neuropsychologia.2011.08.007>
- Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience*, *219*, 111–119. <http://doi.org/10.1016/j.neuroscience.2012.05.042>
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382. <http://doi.org/10.1121/1.2178720>

- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, 25(12), 1649–1653. <https://doi.org/10.1016/J.CUB.2015.04.049>
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674. <https://doi.org/10.1038/nn1082>
- Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, 2(June), 142. <http://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. <http://doi.org/10.1016/j.heares.2013.08.011>
- Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8(May), 57. <http://doi.org/10.3389/fnsys.2014.00057>
- Patel, A. D., & Morgan, E. (2016). Exploring Cognitive Relations Between Prediction in Language and Music. *Cognitive Science*, 1–18. <http://doi.org/10.1111/cogs.12411>
- Phillips-Silver, J., Toiviainen, P., Gosselin, N., Turgeon, C., Lepore, F., & Peretz, I. (2015). Cochlear implant users move in time to the beat of drum music. *Hearing Research*, 321, 25–34. <http://doi.org/10.1016/j.heares.2014.12.007>
- Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science (New York, N.Y.)*, 308(5727), 1430. <http://doi.org/10.1126/science.1110922>
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2-3), 203-228. <http://doi.org/10.1007/s11168-006-9004-0>

- Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, 18(9), 451–456. <http://doi.org/10.1016/j.tics.2014.05.006>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–10. <http://doi.org/10.1016/j.tics.2006.12.002>
- Pihko, E., Leppäsaari, T., Leppänen, P., Richardson, U., & Lyytinen, H. (1997). Auditory event-related potentials (ERP) reflect temporal changes in speech stimuli. *NeuroReport*, 8(4), 911–914. <http://doi.org/10.1097/00001756-199703030-00019>
- Pike, K. L. (1947). On the Phonemic Status of English Diphthongs. *Language*, 23(2), 151. <http://doi.org/10.2307/410386>
- Pitt, M. a., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology. Human Perception and Performance*, 16(3), 564–573. <http://doi.org/10.1037/0096-1523.16.3.564>
- Ponton, C. W., & Eggermont, J. J. (2001). Of kittens and kids: Altered cortical maturation following profound deafness and cochlear implant use. *Audiology and Neuro-Otology*, 6(6), 363–380. <http://doi.org/10.1159/000046846>
- Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology*, 5(September), 1048. <http://doi.org/10.3389/fpsyg.2014.01048>
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., ... Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology*, 27(1), 121–31. <http://doi.org/10.1037/a0031277>
- Quené, H., & Port, R. F. (2005). Effects of Timing Regularity and Metrical Expectancy on Spoken-Word Perception. *Phonetica*, 62(1), 1–13. <http://doi.org/10.1159/000087222>

- Rammsayer, T., & Altenmüller, E. (2006). Temporal Information Processing in Musicians and Nonmusicians. *Music Perception*, 24(1), 37–48. <http://doi.org/10.1525/mp.2006.24.1.37>
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292. [http://doi.org/10.1016/S0010-0277\(99\)00058-X](http://doi.org/10.1016/S0010-0277(99)00058-X)
- Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science*, 29(2), 200–213. <http://doi.org/10.1016/j.humov.2009.08.002>
- Repp, B. H., & Doggett, R. (2007). Tapping to a Very Slow Beat: A Comparison of Musicians and Nonmusicians. *Music Perception: An Interdisciplinary Journal*, 24(4), 367–376. <http://doi.org/10.1525/mp.2007.24.4.367>
- Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic Bulletin & Review*, 20(3), 403–52. <http://doi.org/10.3758/s13423-012-0371-2>
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. <http://doi.org/10.1016/j.humov.2007.07.002>
- Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, 6(FEB). <http://doi.org/10.3389/fpsyg.2015.00089>
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2), 131-141. [https://doi.org/10.1016/0926-6410\(95\)00038-0](https://doi.org/10.1016/0926-6410(95)00038-0)
- Rochette, F., Moussard, A., & Bigand, E. (2014). Music Lessons Improve Auditory Perceptual and Cognitive Performance in Deaf Children. *Frontiers in Human Neuroscience*, 8(July), 1–9. <http://doi.org/10.3389/fnhum.2014.00488>

- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914. <http://doi.org/10.1002/wcs.78>
- Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 336(1278), 367–73. <http://doi.org/10.1098/rstb.1992.0070>
- Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, 8(1), 5027. <http://doi.org/10.1038/s41598-018-23275-9>
- Sacks, H., Schegloff, E. A., & Jefferson, G. A. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 697–735. <http://dx.doi.org/10.2307/412243>
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349. <http://doi.org/10.1111/1469-8986.00172.x>
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–18. <http://doi.org/10.1016/J.TINS.2008.09.012>
- Schwartz, J.-L., Basirat, A., Mé Nard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25, 336–354. <http://doi.org/10.1016/j.jneuroling.2009.12.004>
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10, 295. Retrieved from <https://doi.org/10.1038/nrn2603>
- Segal, O., Houston, D., & Kishon-Rabin, L. (2016). Auditory Discrimination of Lexical Stress Patterns in Hearing-Impaired Infants with Cochlear Implants Compared with Normal Hearing. *Ear and Hearing*, 37(2), 225–234. <http://doi.org/10.1097/AUD.0000000000000243>

- Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage*, 44(3), 1133–1143. <http://doi.org/10.1016/j.neuroimage.2008.09.045>
- Shamma, S. (2008). On the Emergence and Awareness of Auditory Objects. *PLoS Biology*, 6(6), e155. <http://doi.org/10.1371/journal.pbio.0060155>
- Sharma, A., Campbell, J., & Cardon, G. (2015). Developmental and cross-modal plasticity in deafness: Evidence from the P1 and N1 event related potentials in cochlear implanted children. *International Journal of Psychophysiology*. <http://doi.org/10.1016/j.ijpsycho.2014.04.007>
- Sharma, A., Dorman, M. F., & Spahr, A. J. (2002a). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and Hearing*, 23(6), 532–9. <http://doi.org/10.1097/01.AUD.0000042223.62381.01>
- Sharma, A., Dorman, M. F., & Spahr, A. J. (2002b). Rapid development of cortical auditory evoked potentials after early cochlear implantation. *NeuroReport*, 13(10), 1365–1368. <http://doi.org/10.1097/00001756-200207190-00030>
- Sharma, A., Gilley, P. M., Dorman, M. F., & Baldwin, R. (2007). Deprivation-induced cortical reorganization in children with cochlear implants. *International Journal of Audiology*, 46(9), 494–499. <http://doi.org/10.1080/14992020701524836>
- Sharma, A., Spahr, A., Dorman, M., & Todd, N. W. (2002). Early Cochlear Implantation in Children Allows Normal Development of Central Auditory Pathways. *Annals of Otolaryngology & Laryngology*, 111(5_suppl), 38–41. <http://doi.org/10.1177/00034894021110S508>
- Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain research*, 1121(1), 177–189. <https://doi.org/10.1016/j.brainres.2006.08.123>
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration

- of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, 32(25), 8443–8453. <http://doi.org/10.1523/JNEUROSCI.5069-11.2012>
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. <http://doi.org/10.1073/pnas.1008662107>
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587–10592. <http://doi.org/10.1073/pnas.0903616106>
- Street, R. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, 11(2), 139–169. <http://doi.org/10.1111/j.1468-2958.1984.tb00043.x>
- Striano, T., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies*, 7(2), 251–271. <http://doi.org/10.1075/is.7.2.08str>
- Reed, B. S. (2010). *Analysing conversation: An introduction to prosody*. Macmillan International Higher Education.
- Szczepek Reed, B. (2010b). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics*, 42(4), 1037–1059. <http://doi.org/10.1016/j.pragma.2009.09.002>
- ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences of the United States of America*, 112(52), 15833–7. <http://doi.org/10.1073/pnas.1517519112>
- Thaut, M. H., McIntosh, G. C., & Hoemberg, V. (2015). Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. *Frontiers in Psychology*, 5, 1185. <http://doi.org/10.3389/fpsyg.2014.01185>
- Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two

mechanisms of perceptual reactivation. *Cortex*, 77, 1–12.
<https://doi.org/10.1016/J.CORTEX.2016.01.002>

Tierney, A., & Kraus, N. (2013). Music Training for the Development of Reading Skills. *Progress in Brain Research*, 207, 209–241. <https://doi.org/10.1016/B978-0-444-63327-9.00008-4>

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, 8(November), 1–9. <http://doi.org/10.3389/fnhum.2014.00949>

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, 70(5), 345–358. <http://doi.org/10.1007/s00426-005-0222-0>

Timm, L., Agrawal, D., C. Viola, F., Sandmann, P., Debener, S., Büchner, A., ... Wittfoth, M. (2012). Temporal Feature Perception in Cochlear Implant Users. *PLoS ONE*, 7(9), e45375. <http://doi.org/10.1371/journal.pone.0045375>

Toe, D. M., & Paatsch, L. E. (2013). The conversational skills of school-aged children with cochlear implants. *Cochlear Implants International*, 14(2), 67–79. <http://doi.org/10.1179/1754762812Y.0000000002>

Torkildsen, J. von K., Arciuli, J., Haukedal, C. L., & Wie, O. B. (2018). Does a lack of auditory experience affect sequential learning? *Cognition*, 170, 123–129. <http://doi.org/10.1016/j.cognition.2017.09.017>

Torppa, R., Faulkner, A., Huotilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children: the role of auditory working memory and musical activities. *International Journal of Audiology*, 53(3), 182–91. <http://doi.org/10.3109/14992027.2013.872302>

Torppa, R., Huotilainen, M., Leminen, M., Lipsanen, J., & Tervaniemi, M. (2014). Interplay between singing and cortical processing of music : A longitudinal study in children with cochlear implants Interplay between singing and cortical processing of music : A

- longitudinal study in children with cochlear implants. *Frontiers in Psychology*, 5(December), 1–16. <http://doi.org/10.3389/fpsyg.2014.01389>
- Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition*, 100(1), 73–99. <http://doi.org/10.1016/j.cognition.2005.11.006>
- Truy, E., Lina-Granade, G., Jonas, A.-M., Martinon, G., Maison, S., Girard, J., ... Morgon, A. (1998). Comprehension of language in congenitally deaf children with and without cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 45(1), 83–89. [http://doi.org/10.1016/S0165-5876\(98\)00097-4](http://doi.org/10.1016/S0165-5876(98)00097-4)
- Tye-Murray, N. (2003). Conversational fluency of children who use cochlear implants. *Ear and Hearing*, 24(1 Suppl), 82S–89S. <http://doi.org/10.1097/01.AUD.0000051691.33869.EC>
- Vaissière, J. (1991). Rhythm, accentuation and final lengthening in French. In *Music, Language, Speech and Brain. Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5-8 September 1990* (pp. 108–120). http://doi.org/10.1007/978-1-349-12670-5_10
- Tierney, A., & Kraus, N. (2013). Music Training for the Development of Reading Skills. *Progress in Brain Research*, 207, 209–241. <https://doi.org/10.1016/B978-0-444-63327-9.00008-4>
- van der Steen, M. C. M., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7(June), 253. <http://doi.org/10.3389/fnhum.2013.00253>
- van Wieringen, A., & Wouters, J. (2015). What can we expect of normally-developing children implanted at a young age with respect to their auditory, linguistic and cognitive skills? *Hearing Research*, 322, 171–179. <http://doi.org/10.1016/j.heares.2014.09.002>
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2012). Are You Ready to Jump? Predictive Mechanisms in Interpersonal Coordination. *Journal of Experimental*

Psychology: Human Perception and Performance, 39(1), 48–61.
<http://doi.org/10.1037/a0028066>

Vesper, C., Van Der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3–4), 517–530. <http://doi.org/10.1007/s00221-011-2706-z>

Tierney, A., & Kraus, N. (2013). Music Training for the Development of Reading Skills. *Progress in Brain Research*, 207, 209–241. <https://doi.org/10.1016/B978-0-444-63327-9.00008-4>

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, 24(2), 560–4. <http://doi.org/10.1016/j.neuroimage.2004.08.039>

Wang, Y., Shafto, C. L., & Houston, D. M. (2018). Attention to speech and spoken language development in deaf children with cochlear implants: a 10-year longitudinal study. *Developmental Science*, (March), e12677. <http://doi.org/10.1111/desc.12677>

Wechsler-Kashi, D., Schwartz, R. G., & Cleary, M. (2014). Picture Naming and Verbal Fluency in Children With Cochlear Implants. *Journal of Speech Language and Hearing Research*, 57(5), 1870. http://doi.org/10.1044/2014_JSLHR-L-13-0321

Wilson, M., & Knoblich, G. (2005). The Case for Motor Involvement in Perceiving Conspecifics. *Psychological Bulletin*, 131(3), 460–473. <http://doi.org/10.1037/0033-2909.131.3.460>

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957–68. <http://www.ncbi.nlm.nih.gov/pubmed/16615316>

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neurosci.*, 10, 420–422. <http://dx.doi.org/10.1038/nn1872>

- Tierney, A., & Kraus, N. (2013). Music Training for the Development of Reading Skills. *Progress in Brain Research*, 207, 209–241. <https://doi.org/10.1016/B978-0-444-63327-9.00008-4>
- Zamm, A., Wellman, C., & Palmer, C. (2016). Endogenous rhythms influence interpersonal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 42(5), 611–616. <http://doi.org/10.1037/xhp0000201>
- Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The Impact of Musicianship on the Cortical Mechanisms Related to Separating Speech from Background Noise. *Journal of Cognitive Neuroscience*, 27(5), 1044–1059. http://doi.org/10.1162/jocn_a_00758
- Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107(13), 5768–5773. <http://doi.org/10.1073/pnas.1000121107>
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron*, 77(5), 980–991. <https://doi.org/10.1016/J.NEURON.2012.12.037>
- Zwaan, R. a, & Radvansky, G. a. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185. <http://doi.org/10.1037/0033-2909.123.2.162>