Université de Montréal

Prédictions dans le domaine auditif : études électrophysiologiques

Par

Amour Simal

Département de Psychologie, Faculté des Arts et Sciences

Thèse présentée en vue de l'obtention du doctorat (Ph. D.)

en Psychologie, option Neuroscience Cognitive

Juin 2023

Université de Montréal

Unité académique : Département Psychologie, Faculté des Arts et Sciences

---

Cette thèse intitulée

Prédictions dans le domaine auditif : études électrophysiologiques

Présenté par

Amour Simal

A été évalué(e) par un jury composé des personnes suivantes

Martin Arguin
Président-rapporteur

Pierre Jolicoeur
Directeur de recherche

Robert Zatorre
Codirecteur

Simone Dalla Bella
Membre du jury

Natalie Phillips
Examinateur externe

Résumé

Le but de cette thèse était d'étudier les processus de prédiction dans le domaine auditif et de l'activité cérébrale associée à ces prédictions. L'électroencéphalographie a été utilisée afin de mesurer l'activité électrique du cerveau, ainsi que la technique de potentiels reliés aux évènements (PRE) qui permet de mesurer l'activité lié à des processus d'intérêt à la milliseconde près. La grande majorité des études existantes s'intéressent aux prédictions de manière indirecte, par l'observation de signaux d'erreur ou de confirmation de prédictions. Contrairement à ces études, nous avons développé des paradigmes novateurs qui permettent de créer un contexte dans lequel un ou plusieurs stimuli permettent de générer des prédictions. Dans une première étude, les résultats démontrent qu'un son qui permet la prédiction de sons subséquents est accompagné d'une augmentation de l'amplitude des composantes de PRE auditives, la N1 et plus particulièrement la P2, aux électrodes frontocentrales. Dans une seconde étude, cette modulation était entrainée par une modification marquée des oscillations entre 4 et 7 Hz, dans la bande de fréquence thêta. Une troisième étude, était focalisé sur l'activité de PRE en lien avec des prédictions d'ordre temporel. En créant des patrons rhythmiques contextuels dans l'expérience, nous avons pu observer qu'un son qui permet d'identifier le patron, et donc d'anticiper la suite, génère une positivité précoce, et ce, même si le patron rhythmique est non-pertinent à l'exécution de la tâche. Ces études sont une démonstration de faisabilité et fondent une base solide pour la poursuite de la recherche sur les processus dynamiques liés à la génération de prédictions.

Mots-clés : Électrophysiologie, Audition, Prédictions, Potentiels reliés aux évènements, Oscillations

Abstract

The goal of this thesis was to study predictive processes in the auditory domain, and to find the electrophysiological signature associated with those processes. We used electroencephalography to measure the brain electrical activity, as well as the event-related potentials technique (ERP), allowing us to measure brain activity of interest with a millisecond precision. The majority of existing studies that look at predictive processes indirectly, by measuring prediction error or prediction confirmation. Contrary to those studies, we developed novel paradigms that allowed us to generate contexts in which one or more specific auditory stimuli allow the generation of predictions. In a first study, we showed a tone allowing the prediction of other tones to be presented elicits larger N1 and P2 auditory ERP components at frontocentral electrodes. In a second study, we showed this modulation is driven by an increase in oscillatory activity in the theta frequency band, between 4 and 7 Hz. In the third study of this thesis, we were interested in ERP activity related to temporal predictions. By creating contextual rhythmic patterns, we were able to determine that a tone allowing identification of the currently heard pattern, that would allow prediction of when other tones are to be heard, generate and early positivity, even though the rhythmic pattern was task irrelevant. These studies are proof of concept and a solid basis for future research on the dynamics related to predictions in the brain.

Keywords : Electrophysiology, Auditory, Predictions, Event-related potentials, Oscillations

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

ACI : Analyse en Composantes Indépendantes

APA : American Psychological Association

EEG : Électroencéphalographie

ERP : *Event-related potentials* (voir PRE en français)

Hz : Hertz

ICA : *Independant Component Analysis*

ISI : Intervalle interstimuli/*Interstimuli interval*

MEG : Magnétoencéphalographie

MMN : *mismatch negativity*

PRE : Potentiels reliés aux évènements

RP : *Repetition positivity*

SNR : *Signal to noise ratio*

Remerciements

Je souhaite tout d'abord remercier mon directeur de recherche, Pierre Jolicoeur, qui m'a suivi et supporté durant l'entièreté de mes études graduées. Tu as été un modèle, un mentor et un ami durant toutes ces années. Tes encouragements et ta confiance en moi m'ont permis de me trouver et de me développer en tant que chercheuse. Je n'aurais pas pu rêver meilleur mentor. J'espère un jour être pour d'autres étudiants le support que tu as été pour moi.

Je souhaite ensuite remercier mon co-directeur, Robert Zatorre, qui m'a accueilli chaleureusement dans son laboratoire. Tes conseils et ton support ont été inestimables. J'ai appris beaucoup dans ton laboratoire et cela m'a permis d'étendre mon champ de connaissances.

Je remercie également ma famille, qui ont été d'un support inébranlable dans ces années pas toujours faciles et qui ne m'ont pas laissé abandonner. Ma mère, Anita, qui a été là physiquement pour m'aider quand j'en avait besoin et qui m'a encouragée durant toutes les étapes. Mon père, Jean Luc, qui a toujours été à l'écoute de mes succès, de mes échecs, et de mes doutes. Mon frère, David, à qui j'ai toujours pu me confier.

Merci à mes amies et amis, toutes et tous, vous avez marqué mon parcours et avez été indispensables à ma réussite. Remerciement spécial à Brandi qui a été présente durant mon parcours et qui est mon modèle, mon amie et la preuve qu'on peut réussir. Martine, Marie-Ève, Mathieu, mes collègues durant le début de mes études graduées. Alix, Arielle, mes amies d'après fin du monde. À Marie-Audrey et Catherine, merci d'être là pour moi et de m'avoir aidé à finir mon powerpoint pour ma défense de thèse. À Rosemarie, qui m'a apporté confiance en moi et une nouvelle passion libératrice dans le dernier segment de mon doctorat.

À mes animaux qui ne liront jamais ces lignes mais qui ont été un support émotionnel important : Capuchon, Napoléon, Charlotte et Mousse.

Contribution des Auteurs

La section suivante résume les contributions des auteurs aux trois articles scientifiques formant la présente thèse

Chapitre 2 : Signal informativeness for sequence structure modulates human auditory cortical responses

Le recrutement des participants et la collecte des données avaient été effectuées par FV et CL, avec l'assistance d'étudiants du laboratoire de PJ. Les buts et hypothèses ont été développés par AS et PJ, avec la contribution de FV, CL et PB. Les analyses, l'interprétation des données, l'écriture du manuscrit et sa soumission ont été réalisées par AS. PJ, FV et PB ont contribué à relecture du manuscrit et au processus de révision.

Chapitre 3 : Oscillations linked to auditory informativeness and context disambiguation

Le recrutement des participants et la collecte des données avaient été effectuées par FV et CL, avec l'assistance d'étudiants du laboratoire de PJ. Les buts et hypothèses ont été développés par AS. Les analyses, l'interprétation des données et l'écriture du manuscrit ont été réalisées par AS avec la contribution de RZ et PJ

Chapitre 4 : Detecting auditory temporal regularities: electrophysiological index of tracking and identification of disambiguating information

Le recrutement des participants et la collecte des données ont été effectuées par AS, avec l'assistance d'étudiants stagiaires du laboratoire de PJ. Les buts et hypothèses ont été développés par AS. Les analyses, l'interprétation des données et l'écriture du manuscrit ont été réalisées par AS avec la contribution de RZ et PJ

Chapitre 1 – Introduction

1.1 Objectif de la thèse

L'objectif de cette thèse est d'étudier les réponses cérébrales dans le contexte du codage prédictif dans le domaine auditif. À l'aide de l'électroencéphalographie (EEG), il est possible d'étudier les processus dynamiques, la résolution temporelle étant de l'ordre de la milliseconde. À l'aide d'un nouveau type de paradigme, j'ai examiné les signatures électrophysiologiques en lien avec l'identification de régularités dans le contexte et la génération de prédictions.

Dans une première étude (Chapitre 2), je montrerai un indice lié à la génération de prédictions que plus de sons seront présentés dans la séquence. À l'aide d'Analyse en Composantes Indépendantes, le signal sera examiné plus en détail pour montrer différents patrons d'activation d'électrodes qui contribuent au signal de prédictions. Les données des expériences du Chapitre 2 seront également analysées dans le domaine fréquentiel, et présentées dans le Chapitre 3, afin d'examiner les modulation des oscillations qui peuvent être des indices de communication à longue distance dans le cerveau, ou encore des processus d'activation et d'inhibition. Dans une deuxième étude (Chapitre 4), je démontrerai un indice électrophysiologique associé à des prédictions quant au moment de présentation de sons subséquents.

1.2 Contexte théorique

"The whole function of the brain is summed up in: error-correction."
W. Ross Ashby[1]

Notre environnement est complexe et traiter l'entièreté de l'information sensorielle à chaque moment est tout simplement impossible. Heureusement, notre cerveau nous permet de créer des représentations stables et de sélectionner les informations pertinentes par des mécanismes attentionnels. En plus d'avoir cette capacité de conserver la représentation de l'environnement en mémoire, notre cerveau nous permet d'anticiper les évènements. Ceci, faisant allusion aux prédictions, permet de diriger l'attention vers les stimuli d'intérêt avant même leur présentation, rendant ainsi leur perception plus complète, rapide et efficace. Les prédictions sont diverses et quelques exemples sont

---

[1] Aphorisme non-daté. http://www.rossashby.info/index.html

des prédictions quant à la continuité de la stabilité de l'environnement (les objets sur mon bureau ne disparaitront pas de ma vision périphérique sans intervention externe), des prédiction sur des changements auto-initiés (si j'appuye sur des touches de mon clavier, du texte va s'ajouter sur mon écran, ou encore, si je tape mes mains ensemble, je vais entendre un bruit) ou encore des changements externes (j'ai vu un éclair, je vais entendre un coup de tonnerre). Le fait d'avoir ces mécanismes attentionnels et prédictifs est directement utile à la survie, et les dysfonctionnements peuvent être extrêmement handicapants.

Le déploiement de l'attention peut se faire de deux manières principales. Soit par des processus ascendants, ce qui se produit lorsque notre attention est attirée par un stimulus saillant (une lumière soudaine, un bruit fort), soit par des processus descendant, lorsque nous cherchons un objet ayant des caractéristiques spécifiques (Katsuki & Constantinidis, 2014). Par exemple, si nous voyons un éclair dans le ciel, notre attention sera capturée de manière ascendante. L'attente d'entendre le tonnerre sera ensuite générée de manière descendante et nous permettra de nous y préparer et d'y porter attention (si l'on souhaite mesurer la distance entre l'éclair et nous, par exemple), ou de l'ignorer afin de ne pas être distraits dans nos tâches. Dans ce contexte, les prédictions sont utiles dans le déploiement attentionnel descendant, puisqu'elles permettent de générer des hypothèses quant à la localisation, l'identité ou encore le moment de présentation du stimulus cible. Une des premières preuves de ceci a été amenée par Posner (1980) avec son paradigme d'indiçage. Dans ce type de paradigme, un indice prédit, avec plus ou moins de précision et de fiabilité, l'endroit auquel sera présenté un stimulus visuel d'intérêt. Lorsque l'indice est valide, et permet de prédire adéquatement l'endroit de présentation du stimulus, celui-ci est détecté plus rapidement et avec plus de précision (Prinzmetal et al., 2005). Dans le domaine auditif, les mêmes types d'avantages comportementaux ont été décrits. Lange (2009) a montré, en modulant la prévisibilité de la tonalité et/ou le moment de présentation d'un son, que la capacité d'effectuer des prédictions menait à des réponses plus rapide que quand les stimuli étaient imprévisibles.

Dans le domaine auditif, les prédictions semblent être d'une importance particulière, étant donné la nature éphémère des sons et de leur déroulement dans le temps. Il est important de pouvoir évaluer, encoder, et traiter les stimuli le plus efficacement possible sans manquer de l'information. Ainsi, l'extraction de régularités dans l'environnement permet de déduire des règles qui nous permettront de déterminer rapidement quels stimuli nécessitent notre attention. Différentes sources de prédictions ont été identifiées (voir Bendixen et al., 2012). Les prédictions peuvent provenir de l'information auditive contenue dans la séquence actuelle, par exemple patron se répétant note haute suivie de note basse, ou encore provenir d'information extraite de la mémoire à long terme, telle que les premières notes d'une chanson connue entraînent sa reconnaissance et la capacité de prédire la suite. Les prédictions proviennent parfois de l'information obtenue autrement que par des liens audition-audition, comme l'association entre des stimuli visuels et auditifs peut permettre les prédiction (l'éclair et le tonnerre), ou encore les actions motrices (applaudir, le clic d'une souris d'ordinateur). Peu importe les prédictions, elles nécessitent un apprentissage des régularités. Dans le quotidien, cet apprentissage ne se fait habituellement pas consciemment mais relève plutôt d'un apprentissage statistique (Turk-Browne, 2012). Le concept d'apprentissage statistique est venu en premier des études sur le langage en psychologie développementale (Saffran et al., 1996). Dans le langage parlé, il n'y a, a priori, pas d'indice quant au commencement et à la fin des mots. Si nous entendons quelqu'un parler une langue qui nous est inconnue, il nous est extrêmement difficile de segmenter mentalement ce que nous entendons, alors qu'il s'agit d'un exercice trivial pour notre langue maternelle. Dans leur expérience, Saffran et al. (1996) présentaient des assemblages de trois syllabes (non-mots de trois syllabes les uns après les autres sans interruption à des enfants. L'idée était que des syllabes se suivant avec une haute probabilité sont plus susceptibles de faire partie du même mot que des syllabes se suivant avec une faible probabilité. Après deux minutes d'exposition, les enfants étaient en mesure de différencier les nouveaux mots de syllabes assemblées aléatoirement. Ce type d'apprentissage est présent dans les modalités visuelles, tactiles et auditives et pour différents types d'apprentissages.

Indices électrophysiologiques

Les prédictions se font de manière rapide et transitoire. Pour cette raison, les mesures électrophysiologiques sont idéales pour mesurer ce type de processus. L'EEG possède une excellente résolution temporelle, de l'ordre de la milliseconde, et permet de mesurer, de manière non invasive, l'activité sous-jacente à des processus sensoriels, cognitifs et moteurs (Luck et Kappenman, 2011). La technique de potentiels reliés aux évènements (PRE) permet de mettre en évidence ces processus en augmentant le rapport signal/bruit par le moyennage d'un grand nombre de segments EEG, ce qui a pour effet de réduire l'activité aléatoire et de ne conserver que l'activité systématique se produisant autour de la présentation d'un stimulus ou d'une activité motrice (Luck, 2014). Dans les études en audition, les composantes principales de PRE étudiées sont la N1, se produisant environ 100 ms après la présentation d'un stimulus auditif, et la P2, se produisant environ 200 ms après la présentation du stimulus, toutes deux mesurables aux électrodes frontocentrales. Une composante plus précoce, la P1 ou P50, se produisant 50 ms environ après la présentation du stimulus, est également visible, bien que moins étudiée (pour une revue succincte des composantes de PRE auditives, voir Winkler et al., 2013). Il existe de nombreuses études s'intéressant aux réponses cérébrales lors de processus de prédiction dans le domaine auditif. La plupart de ces études ont montré l'existence de processus prédictifs par mesures indirectes de stimuli correspondant à des prédictions ou ne correspondant pas (pour une revue voir Bendixen et al., 2012), plutôt que d'observer l'activité cérébrale lorsque des prédictions sont effectuées.

D'abord, mentionnons les paradigmes existants qui s'intéressent aux réponses électrophysiologiques à un stimulus violant la prédiction (*mismatch*), qui sont prédominants dans la littérature sur les prédictions dans le domaine audifit. La composante de PRE principalement observé dans ce type de paradigme est la *mismatch negativity* (MMN). La MMN a été découverte par Näätänen et al. (1978) dans une tâche visant à observer les différences d'activité PRE entre des stimuli auditifs cibles (rares, entre 0 et 9% des sons présentés), vers lesquels le déploiement attentionnel est nécessaire, et des stimuli standards, qui ne sont pas pertinents à la tâche. Une onde négative, superposée aux PREs auditifs, était observée lorsque les stimuli rares étaient présentés et que l'attention était dirigée vers ceux-ci. Cette négativité, observable environ 150 à 250 ms après la présentation auditive, a plus tard été séparée de la N2 auditive et décrite comme la MMN (Näätänen et al., 1982; Näätänen et Gaillard, 1983). Les études subséquentes ont démontré que cette négativité était générée lorsque des stimuli déviants étaient présentés parmi des standards (paradigme *oddball*), et ce, même si l'attention n'était pas dirigée vers ceux-ci. Ainsi la MMN est générée automatiquement par le cerveau en réponse à des changements dans la stimulation auditive. Cette composante, décrite dans près de 2000 publications (pour des revues de littérature, voir, entre autres, May et Tiitinen, 2010; Näätänen et al., 2007; Winkler, 2007). Elle est traditionnellement associée à des violations de prédictions induites par la constance de stimuli auditifs. Elle est maximale aux électrodes frontocentrales (avec une référence au nez) avec une contrepartie positive aux électrodes postérolatérales et est générée, au moins partiellement, dans les aires auditives du cortex supratemporal (Giard et al., 1990; Molholm et al., 2005; Rosburg et al., 2005; Sabri et al., 2006; Shalgi et Deouell, 2007).

Dans les paradigmes « *oddball* » standards, une explication de la MMN vient d'un modèle basé sur l'adaptation (May & Tiitinen, 2010). En effet, la présentation répétée d'une stimulus « standard » provoque une baisse de réponse et d'amplitude des PRE par des processus d'adaptation neuronale et d'inhibition latérale. Ainsi un stimulus « déviant » génère, en comparaison une activation plus grande, puisque les neurones à fréquence préférentielles ne sont pas adaptés et ceci est reflété dans la MMN (May & Tiitinen, 2004; May & Tiitinen, 2001). L'autre théorie principale de la source de la MMN provient d'un modèle basé sur la mémoire, qui assume que la régularité est retenue et que chaque entrée est comparée avec l'attente. Dans ce modèle la MMN est une erreur de prédiction. Depuis les études sur les paradigmes *oddball* simple, plusieurs résultats d'expériences favorisent le modèle basé sur la mémoire et sur les prédictions. Un de ces résultats provient du fait que le modèle sur l'adaptation assume que la MMN est une modulation de la N1 auditive, alors qu'il a été démontré que les générateurs sont différents (May & Tiitinen, 2004, 2010). En effet la MMN a deux sous-composantes (Garrido et al., 2009; Näätänen & Kreegipuu, 2012). La composante supratemporale bilatérale, dans le gyrus temporal supérieur, est impliquée dans la perception et la détection du changement, ainsi que dans les processus mnésiques. La composante frontale, prédominante dans l'hémisphère droit, dans le gyrus frontal inférieur, a un rôle de comparateur et est liée à l'initiation de changement attentionnel. Un autre argument en faveur de la MMN basée sur les erreurs de prédiction est que des régularités complexes sont capables de générer une MMN. Bendixen et al. (2008) ont utilisé une régularité selon laquelle les sons d'une durée longue sont suivi d'un son de haute fréquence (long ou court) et qu'un son court est suivi d'un son de basse fréquence, ou vice versa, et observaient une MMN pour les sons ne respectant pas la règle (les erreurs de prédiction). Dans ce cas, l'apparition d'une MMN ne peut pas être simplement causée par l'adaptation de neurones ayant une fréquence préférentielle.

Une autre question dans ces études s'intéressant à la violation de régularités par le biais de la MMN, est de déterminer si les réponses se font de manière prospective ou rétrospective. La théorie prospective (prédictive) explique la MMN par le fait que le système auditif extrait des régularités de la stimulation, puis compare activement chaque entrée dans le système aux prédictions dérivées de ces régularités (Winkler, 2007). Pour une analogie, cela serait comme comparer deux documents l'un à côté de l'autre, un d'eux étant la prédiction qui a été générée, l'autre étant le nouveau document. Il est toutefois également possible d'expliquer la MMN par une vue rétrospective dans laquelle le cerveau tente de faire correspondre le stimulus présent aux stimuli le précédant seulement lors de sa présentation. Dans l'analogie, ça correspondrait au fait de lire le nouveau document et d'ouvrir un livre à chaque mot lu. Une étude de Grimm et al. (2011) a apporté des preuves en faveur d'un modèle prédictif. Dans un paradigme oddball standard, ils ont, en plus de mesurer les réponses à latence longue (la MMN), mesuré les PREs à latence moyenne se produisant 10 à 50 ms après la présentation des stimuli (la P0, Na, Pa, Nb). Les sons déviants généraient une MMN typique, mais également une composante Nb (environ 40 ms après le début de présentation du son, Picton et al., 1974) de plus grande amplitude. Une possibilité est que cette modulation se produit par un effet réfractaire de neurones répondant à une fréquence spécifique. Toutefois, une condition contrôle dans leur expérience a montré que cette modulation ne se produisait pas lorsqu'un son n'était pas un déviant dans le contexte, mais avait une fréquence différente des sons le précédant. La courte latence de cette modulation étant peu facilement explicable par une réponse rétrospective, les auteurs argumentent en faveur d'un mécanisme prédictif.

Une autre étude effectuée par Bendixen et al. (2009) a également apporté des arguments pour départager si les signaux de mismatch résultent d'une prédiction ou d'un processus rétrospectif. Pour ceci, ils ont utilisé des séquences de sons contenant une omission. Le stimulus auditif omis pouvait être prédit avant (condition prédictible) ou après (condition restaurable) son omission. Une condition contrôle dans laquelle l'identité du son ne pouvait être récupérée était également utilisées. Toutes les séquences présentées contenaient le même nombre de sons ayant la même durée et le même intervalle interstimulus. Ces conditions étaient créées en présentant deux sons de la même fréquence séquentiellement. Pour la condition prédictible, le deuxième son de même fréquence était omis et son identité était connue puisqu'identique au premier son, tandis que pour la condition restaurable, le premier son de même fréquence était omis, et son identité pouvait être connue rétrospectivement par la présentation du deuxième son. Dans la condition contrôle, la fréquence des sons était complètement aléatoire. En observant l'activité de PRE au son omis, ils ont trouvé une activité positive, entre 10 et 50 ms après le début théorique du son omis, mesurée seulement pour la condition prédictible. Ils expliquent la différence très précoce dans l'activité PRE en argumentant que le cerveau se prépare à traiter un son spécifique avant sa présentation et que ce traitement n'est arrêté que lorsque l'omission est détectée, ce qui peut demander quelques dizaines de ms. Cette expérience a permis aux auteurs d'apporter une preuve que le cerveau effectue des prédictions lorsque le contexte le permet.

D'autres composantes de PRE ont également été décrites comme répondant à une violation des prédictions. La négativité précoce antérieure droite (*early right anterior negativity, ERAN*; Koelsch et al., 2000) est observée lorsque l'attente concernant la syntaxe musicale est violée. N'étant pas induite par une déviation physique des stimuli par rapport à un standard, mais plutôt par la présentation d'accords harmoniquement inappropriés, la ERAN ne peut pas être assimilée à une MMN. De plus, contrairement à la MMN, l'amplitude de la ERAN est modulée par le degré de convenance harmonique

(Koelsch et al., 2001). Dans le langage, des composantes de PRE similaires ont été décrites. La négativité précoce antérieure gauche (*early left anterior negativity, ELAN*; Hahne et Friederici, 1999) est l'équivalent de la ERAN pour le langage et reflète les violations de règles syntaxiques. La ERAN et la ELAN se produisent environ 150 à 200 ms après la présentation des stimuli et ont toutes deux une distribution antérieure. Toutefois, la ELAN est latéralisée vers la gauche, du côté vers lequel le langage est latéralisé chez la plupart des gens (Berker et al., 1986), tandis que la ERAN est latéralisée à droite, au-dessus des aires auditives plus spécialisées dans la résolution spectrale impliquée dans la perception musicale (Zatorre et al., 2002). D'autres composantes plus tardives liées au langage ont également été décrites. La N400, se produit 250 à 400 ms après une violation d'attente sémantique (Kutas et Hillyard, 1980), tandis que la P600 se produit environ 600 ms après une violation de la structure des phrases (Friederici, 2002).

Un cas extrême de violation des prédictions est quand un stimulus attendu n'est tout simplement pas présenté. Dans ce cas de figure, les études ont démontré que le cerveau répondait à ces omissions et que le cortex auditif générait une activité de type MMN, appelée *omission MMN* (Raij et al., 1997; Yabe et al., 1997, 1998). Ceci implique que le cortex auditif ne nécessite pas d'entrée sensorielle pour détecter une violation de prédiction. De plus, une étude de Bendixen et al. (2009), décrite plus haut, a montré qu'un son prévisible, mais omis, générait tout de même une activité cérébrale précoce, dans les 50 premières millisecondes, similaire à un son réellement présenté. Ceci indique, selon les auteurs, que le cerveau est préparé à traiter le son, et que ce traitement n'est interrompu que quand l'omission est détectée, après un délai.

D'autres paradigmes se sont intéressés à étudier les stimuli se conformant aux prédictions (*match*). Haenschel et al. (2005) ont été un des premiers à décrire l'activité électrophysiologique en lien avec la présentation d'un son hautement prévisible. Dans leur paradigme de *roving standard* (standard changeant), un stimulus auditif est répété et devient le standard, de plus en plus prévisibles. Après un certain nombre de répétitions (deux, six, ou trente-six), un son différent (déviant) est présenté et devient le nouveau standard. Ainsi, en étudiant la modulation de réponse selon le nombre de répétitions, ils ont pu mesurer une positivité se superposant à l'activité de PRE, appelée *repetition positivity* (RP). Cette positivité était mise en évidence en comparant les PREs générés par un son répété, hautement prévisible, aux PREs générés par un son contrôle (tel qu'un des premiers sons présentés). Elle était mesurable de 50 à 250 ms après le début de présentation d'un son se répétant, aux électrodes frontocentrales, et créait ainsi une atténuation visible de la N1. Aussi, son amplitude augmentant avec le nombre de répétitions, Haenschel et al. (2005) l'ont interprétée comme un corrélat d'une adaptation rapide sous-jacent à la formation de représentations en mémoire acoustique. Ils ont également suggéré que le niveau de traitement d'un son est diminué plus celui-ci est répété, et ce, parce que son identité devient prévisible. Une des critiques principales de ce type d'expérience est que cette positivité pourrait être simplement causée par un phénomène d'habituation et d'effet réfractaire des neurones ayant une fréquence préférentielle liée à la présentation répétée d'un même son. Toutefois, une étude de Schadow et al. (2009) a montré l'existence d'une RP avec des régularités plus complexes. Ils ont utilisé des séquences caractérisées par une augmentation ou une diminution graduelle de ton, d'un son au suivant. Puisque les sons ne sont pas prévisibles en étant identiques au son précédant mais en suivant le patron descendant ou ascendant, leurs résultats d'analyse de PRE montraient une RP, suggérant que cette composante n'est pas uniquement due à des processus de bas niveau, tel qu'un effet réfractaire des neurones ayant une préférence de fréquence. Cette RP se produit également avec des régularités encore plus complexes. Bendixen et al. (2008) ont utilisé des séquences contenant des sons longs ou courts, aigus ou graves. La régularité était créée en générant des séquences dans lesquelles les sons courts étaient suivis d'un son aigu, et les sons longs d'un son grave, ou vice versa. Ainsi, un son d'une identité spécifique pouvait être un standard (respecter la régularité) ou un déviant selon le contexte l'entourant. Dans une écoute passive, sans informer les participants des régularités, les standards généraient une RP, tandis que les déviants généraient une MMN, tel qu'attendu. Étonnamment, bien que 10 participants sur 14 avaient perçu une possible régularité, aucun d'entre eux n'était capable de la décrire. Ceci indique que le cerveau est capable d'extraire des régularités et de générer un patron de prédictions sans que la perception consciente soit nécessaire.

Les études décrites plus haut se sont intéressées aux réponses électrophysiologiques à un son selon qu'il était prédit (ou attendu) ou pas. Ces études s'intéressent aux différences entre un son conforme à la prédiction et un son non-conforme à la prédiction. Dans cette thèse, je souhaite plutôt observer l'activité cérébrale lorsqu'une prédiction se fait, c'est-à-dire quand la présentation d'un son permet d'inférer les caractéristiques de sons *subséquents*. Aussi, les paradigmes seront développés afin de réduire le risque que les effets observés soient dus à des processus de bas niveau tel que l'habituation.

1.3 Buts et hypothèses

Dans cette thèse, je présente des paradigmes novateurs développés dans le but d'étudier l'activité cérébrale lorsqu'une prédiction peut être faite. Ceci a été fait en créant un nombre restreint de régularités dans des séquences auditives qui sont contextuelles aux expériences. Dans ces séquences un ou plusieurs sons sont informatifs du type de séquence entendue. Le son permet donc l'identification de la régularité et la génération de prédictions.

La première étude portait sur les prédictions que d'autres sons seront présentés. Les régularités créées provenaient du fait que les séquences de sons pouvaient contenir un, trois, ou cinq sons. Les participants effectuaient une tâche de mémoire dans laquelle ils entendaient deux séquences contenant le même nombre de sons séparées par un intervalle de rétention silencieux et ils devaient indiquer si la hauteur des sons étaient identique entre les séquences. Pour un groupe de participants, chaque essai contenait aléatoirement des séquences de un, trois ou cinq sons (toujours le même nombre pour les deux séquences). Dans ce cas, lors de la présentation de la première séquence, nous nous attendions à ce que le deuxième et le quatrième son, s'ils sont présentés, soient prédicteurs de la présentation d'au moins un autre son, les séquences ne pouvant pas ne contenir que deux ou quatre sons. Par contre, dans la deuxième séquence, qui contient le même nombre de sons que la première, information connue avant sa présentation, le deuxième et le quatrième son, si présentés, n'apportent pas d'informations supplémentaires. Un autre groupe de participants a effectué la même expérience à l'exception clé que les trois types de séquences étaient présentées en bloc, donc le nombre de sons contenus dans chaque séquence (même la première) était connu avant le début de l'essai et le deuxième et le quatrième son, si présentés, n'apportent pas d'information. Nous avons observé l'activité EEG et les PRE en lien avec les sons informatifs.

La seconde étude portait sur les mêmes données que la première étude, mais nous étions intéressés à comprendre les mécanismes oscillatoires en lien avec les modulations électrophysiologiques liées à la résolution des ambigüités du contexte. Nous avons utilisé les ondelettes de Morlet, ainsi que la transformée de Hilbert afin de déterminer les fréquences impliquées dans la génération de prédictions et d'émettre des hypothèses d'ordre fonctionnel.

Dans la troisième étude, nous souhaitions observer les prédictions d'ordre temporel. Nous avons créé trois types de séquences caractérisées par l'intervalle interstimulus (ISI) entre les sons, soit trois patrons rhythmiques. Dans ce contexte expérimental, le deuxième son de la séquence était celui qui permettait de résoudre les ambigüités de quel patron rhythmique était entendu. Soit le deuxième son était présenté après un court ISI et le patron était, court-long-court-long, soit il était présenté après un long ISI et le patron était long-court-long-court, soit il était présenté après un ISI moyen et la séquence était isochrone (ISI tous moyens). De manière similaire à la première étude, nous avons utilisé une tâche de mémoire portant sur la hauteur des sons avec un groupe de participant qui entendait un patron aléatoire à chaque essais (le même pour les deux séquences) et un groupe de participants qui entendait les patrons rhythmiques en bloc. L'hypothèse de cette expérience était que le deuxième son apporte de l'information lors de la présentation de la première séquence uniquement quand le patron rhythmique n'est pas connu à l'avance. Nous avons étudié les prédictions temporelles à l'aide des PRE.

Chapitre 2 – Signal informativeness for sequence structure modulates human auditory cortical responses

Amour Simal[1,2,3], Patrick Bermudez[1,2], Christine Lefebvre[1,2,3], François Vachon[4], & Pierre Jolicœur[1,2,3]

1. Université de Montréal, Montréal
2. BRAMS Laboratory, Montréal
3. Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal
4. École de Psychologie, Université Laval

Published article : Simal, A., Bermudez, P., Lefebvre, C., Vachon, F., & Jolicœur, P. (2021). Signal informativeness for sequence structure modulates human auditory cortical responses. *Psychophysiology*, *58*(3), e13745. https://doi.org/10.1111/psyp.13745

2.1 Abstract

We observed how information about the structure of tone sequences modulates cortical responses in the context of a standard short-term memory (STM) task. Participants heard two sequences of 1, 3, or 5 tones (203 ms on, 203 ms off) interspersed by a silent interval (2 s) and decided whether the sequences were the same or different. In Experiment 1, sequence length was randomized between trials. During the first sequence, the amplitude of the auditory P2 was larger for the second tone in trials with 3 tones, and for the second and fourth tones in trials with 5 tones. We hypothesize the increase in P2 reflected a dynamic disambiguation process because these tones were predictive of a sequence longer than 1 or 3 tones. This hypothesis was supported by the absence of P2 amplitude modulation during the second sequence (when sequence length was known). In Experiment 2, we blocked trials by sequence length to ensure the effects were not caused by some process related to encoding in STM. There was no P2 amplitude modulation in either the first or second sequences. Thus, Tones 2 and 4 had a larger amplitude only when they provided new information about the length of the current tone sequence. To some extent, the auditory N1 also showed those modulations. Independent Component Analysis of the ERPs provided evidence the modulations in P2 amplitude could originate in auditory cortex. These results suggest a rapid dynamic adaptation of auditory cortical responses based on the local informativeness of auditory signals.

2.2 Introduction

The ability to predict or anticipate the occurrence of stimuli and/or events in our environment is a fundamental and key feature of cognition. Through the contribution of prediction to the rapidity and efficiency of cognitive processing, it is directly relevant to survival. The regular unfolding of stimuli over time is a central facet of the world that affords learning or adaptive prediction, and this seems particularly relevant for stimuli in the auditory domain. Rapid, pre-attentive prediction, coding of prediction error, and disambiguation of stimuli seems to be of particular importance. In some circumstances, prediction has been shown to improve both accuracy and speed of responses, in part due to better-targeted allocation of resources (such as attention), whether it is a temporal prediction (Lawrance et al., 2014) or a spatial prediction (Mondor, 1999).

Prediction as been studied in the auditory domain mainly in the processing of language. Indices of brain activity related to such processing have been observed. For example, we can detect event-related potential (ERP) components in response to the violation of phonetic, syntactic, and/or semantic expectations amongst others (Swaab et al., 2012). Less is known, however, about which auditory ERP components may be sensitive to predictive processes during the perception of tone sequences, especially those whose organization is not governed by a musical syntax or simple sensory deviation (e.g., oddball paradigms).

Among the studies having looked at expectation emerging from temporal or tonal regularity in sequences of sounds, it has often been difficult to eliminate the possibility that part of the observed modulation of neural activity arises from simpler low-level phenomena such as habituation. Indeed, Heilbron and Chait (2018) pointed out that the most studied ERP component representing violation of prediction in the auditory domain, the mismatch negativity (MMN), although involving memory, was more easily explained by adaptation than by a predictive coding model (for an extensive review, see May & Tiitinen, 2010; see also Kujala et al., 2007; Näätänen et al., 2007; Winkler, 2007). Other experiments report an attenuation of N1 or P2 amplitude with increasing regularity and a concomitant recovery of those amplitudes, referred to as a ‘repetition positivity’ (RP; Grill-Spector et al., 2006). This RP, reflecting habituation, and sometimes considered a signature of increased processing efficiency, was shown by Costa-Faidella et al. (2011) to be modulated by predictability. Interestingly, Todorovic and de Lange (2012) found that early auditory evoked responses (40–60 ms post stimulus) were attenuated by stimulus repetition, but not expectation, and that, expectation attenuated the intermediate response (100–200 ms), but not repetition. Another group of studies focusing on the role of directed attention in the perception of tone sequences indicate that early auditory components evoked by sounds presented at an expected time are enhanced, showing shorter latencies and/or larger amplitudes (N1 amplitude, Lange et al., 2003; N1 latency, Folyi et al., 2012).

Neither long- nor short-term retrospection has definitive implications for awareness or consciously directed attention (i.e., the process may be pre-attentive or deliberate). Attention is known to modulate auditory components such as the visual and auditory N1, most often augmenting them. Although the stimuli may be attended, the electrophysiological indices of prediction may be 'pre-attentive' in the sense that the predictive process may rely on regularities that are not in consciousness, per se. As has been made clear by a significant accrual of experimental results (see Näätänen et al., 2007; and Bendixen et al., 2012, for reviews), expectations may be generated and conformed to or violated at many different levels of complexity, from simple sensory deviation (as in the simplest oddball paradigm) to syntactic expectations (as in the reinterpretation of sentence structure), to greater complexity still.

In sum, key factors in sorting out these literatures (oddball P300, MMN of various kinds, sensory deviation, rule/expectation violation, etc.) include short- versus long-term memory, and interactions between them, which in turn is somewhat related to rule complexity and attention. Attention is always a factor and often manifests as a potentiation of some components, but this is a separate issue from rule awareness.

Although the auditory N1 and P2 components have long been studied as a dependant complex, and fewer studies have focused on the P2 *per se*, there are reasons to consider them as separable brain responses (ageing, sleep vs. wake studies, and respiratory-evoked potentials; see Crowley & Colrain, 2004). As such, we aimed to use both of these late-latency auditory-evoked components (i.e., the N1 and P2) to examine how the brain reacts to stimulus informativeness and predictability.

Here we describe a pair of studies on auditory short-term memory designed to understand how signal informativeness modulates human auditory cortical responses. In a first experiment, we used a typical short-term memory task consisting of the presentation of a memory set that could contain one, three, or five tones to be remembered (encoding phase). These memory set-size levels (1, 3, or 5) were randomly interspersed in each block of trials. The encoding phase was followed by a silent retention interval (2,000 ms) in which no stimuli were presented (retention phase). A test phase followed during which we presented a tone sequence of the same length as in the encoding phase. The task was to indicate if the test sequence was the same, or different, as the encoded sequence.

Although the original intent of the experiment was to study brain activity during the retention phase, here we focus on brain activity during the encoding and test phases (see Lefebvre et al., 2013, for brain activity related to the retention of auditory representations based on a close variant of the current paradigm, and the following articles for related memory research: Alunni-Menichini et al., 2014; Grimault et al., 2014; Guimond et al., 2011; Lefebvre et al., 2013; Lefebvre & Jolicœur, 2016; Nolden, Bermudez, et al., 2013; Nolden, Grimault, et al., 2013).

Here we focus on contextually-defined sequence effects. Because the sequences had, randomly, one, three, or five tones, we expected the second and fourth tone to contain information about sequence length. The first tone was always expected and did not provide information about how many other tones might occur. A second tone indicated that a third tone was sure to occur (but not necessarily a 4th and 5th). Note that a second tone was presented in 2/3 of the trials, and, as such, it was a relatively high-probability event. When a 2nd and 3rd tone were presented, a fourth and fifth tone occurred on 1/2 of these trials, and none in the other 1/2. That fourth tone, in the same manner as the 2nd tone, contained information about a 5th tone to be heard. If informativeness modulates cortical response, we should see a different activity associated with the presentation of the 2nd and 4th tones, compared to the 1st, 3rd, or 5th tone. We believe we have found evidence supporting this hypothesis by examining the amplitude of the auditory P2 and of the N1, which were larger for the second and fourth tones. These amplitude differences should be seen only during the encoding phase, because sequence length at the time of the test phase was known, and therefore tone positions no longer carried information concerning sequence length.

In a second experiment, we blocked the load conditions, meaning that participants always knew how many tones they would need to encode. Thus, if the effects seen in Experiment 1 were caused by differential informativeness of the tones for the overall sequence length, we should no longer observe any difference between the activity elicited by the 2nd and 4th tone, and the activity elicited by the other tones. If the effects were caused by other, more bottom-up factors which we did not anticipate (or understand), then these effects should also be observed in Experiment 2.

It has sometimes been difficult to distinguish between the auditory P2 and the P3a given both have anterior midline voltage scalp distributions and time courses that are not very different (Polich, 2007; Winkler et al., 2013). Usually, P2 latency would be earlier than P3a latency, but one could imagine situations in which a particularly early P3a could be mistaken for the P2. To address this issue, we used an Independent Component Analysis (ICA; Jung et al., 2001) to decompose the ERPs, in an attempt to tease apart brain activity giving rise to similar, but potentially separable, ERPs. The ICA decomposition included a component with a distribution consistent to a generator in the auditory cortex within the P2 time range that appears to be distinct from the P3a. These results are described in detail in the Results and Discussion sections of the article.

2.3 Materials and methods

2.3.1 Participants

Subjects were primarily recruited from the undergraduate population at Université de Montréal. All reported normal hearing and no history of neurological disorder and gave informed consent to procedures vetted by the university ethics committee. Sample constitution was as follows: 17 participants (12 females, 2 left-handed; average age of 21.5 years (*SD* = 1.9)) for Experiment 1, and 16 participants (10 females, 2 left-handed; average age of 22.8 years (*SD* = 3.6) for Experiment 2. One participated in both experiments. Because this participant's results were similar to that of the others and that their inclusion or exclusion had negligible effects on the grand average ERPs, we included the data in the analyses. Data from two participants from Experiment 1, and one from Experiment 2 were rejected because there were more than 19% of trials rejected due to EEG artifacts (detailed in a subsequent section).

2.3.2 Stimuli

Tones lasted 203 ms (including 20 ms, linear rise and fall times) and were interspersed with 203 ms of silence. They were pure tones corresponding to six notes ranging from C5 to A5 within the C Ionian scale (exact frequencies were: 523 Hz, 587 Hz, 659 Hz, 698 Hz, 784 Hz, and 880 Hz). Tone sequences were comprised of either 1, 3, or 5 tones selected randomly, without replacement, from the available six. Tone were presented, at a comfortable loudness (around 65 dB), using Etymotics ER-1 earphones at a sample rate of 44.1 kHz and a depth of 16 bits.

2.3.3 Procedure

Figure 2.1 illustrates the sequence of events in a trial. Participants heard two sequences of tones, containing the same number of tones (1, 3, or 5), separated by a 2-second silent retention period and were asked to indicate (by button press), after the presentation of a response cue (white circle presented 500 ms after the end of the second sequence), if the second sequence was identical to the first, or different. This delay was included because it was important to observe EEG activity during and immediately after the presentation of the second sequence without overlapping motor activity that would have been associated with an immediate speeded response. We measured RT as the time between the onset of the response cue and the response, but we did not analyze the RTs in any detail because the long forced delay rendered comparisons across conditions uninformative. Set sizes, meaning the number of tones presented in each set of a given trial are later referred to as Load 1, Load 3, and Load 5 conditions. In half of the trials, they heard the first set only in the left ear, and in the right ear on the other half. The second sequence was presented to both ears. When the two sequences differed, two randomly-selected adjacent tones in the second (test) sequence were presented in reverse order, or, for the Load 1 condition, the tone was simply different. Crucially, in the first of the two experiments (mixed experiment), the 3 load conditions (sequence lengths of 1, 3, or 5 tones) were randomly intermixed in each block of 60 trials (20 for each load). There were 10 blocks of 60 trials, yielding a total of 600 trials, equally distributed across load conditions and ear of presentation (half to the left, half to the right). In half of the trials for each load condition, the two sets matched, and they mismatched in the other half of the trials. In the second (or blocked) experiment, all parameters were identical except that trials of a given load were blocked in five consecutive blocks consisting each of 40 trials (total per sequence length of 200). The order of presentation of the different loads was counterbalanced across participants to eliminate training effects in such that load conditions would be presented equally often in each possible position in the test order (1-3-5, 1-5-3, 3-1-5, 3-5-1, 5-1-3, 5-3-1). Participants were not informed of the different memory load conditions, but had one practice block in the mixed experiment, and three practice block (one for each load condition) in the blocked experiment.

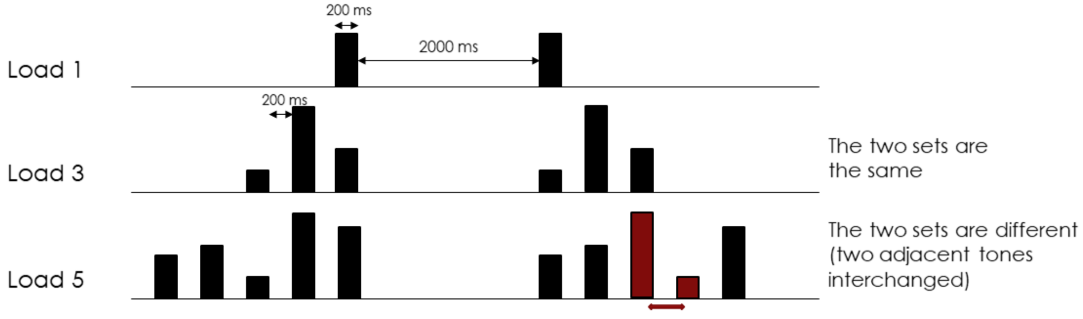

Figure 2.1 Illustration of the paradigm. The bars represent tones, with the height representing pitch. Each trial had three phases: encoding, retention, test. One, three, or five tones were presented at encoding and test (always the same number in a given trial). Tones had a duration of 200 ms, and silences between them were also 200 ms in length. The retention phase lasted for 2000 ms. In Experiment 1, load conditions were mixed randomly within blocks; in Experiment 2, they were constant/fixed within blocks. The encoding and test sequences could be the same or different. When different, the order of two adjacent tones in the test was reversed. See Methods sections for further details.

2.3.4 EEG analyses

EEG signals were sampled continuously at 256 Hz with a 72-channel BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes within an elastic cap with placements that followed the International 10–10 system (Sharbrough et al., 1991). Additional electrodes were placed at the left and right outer canthi to track horizontal eye movements, below the left eye to track vertical eye movements (in combination with the frontal electrode Fp1 that is placed above the left eye), and on the left and right mastoids. EEG data were re-referenced offline to the average of the two mastoids. The difference between the electrodes on the left and right outer canthi was defined as the horizontal electrooculogram (HEOG), and the difference between the electrode below the left eye and Fp1 as the vertical eletrooculogram (VEOG).

EEG data was analyzed using EEGLAB (Delorme & Makeig, 2004), ERPLAB (Lopez-Calderon & Luck, 2014), and custom Matlab code. An Independent Component Analysis (ICA, see

Drisdelle et al., 2017) was performed on the EEG from each subject and components corresponding to ocular artefacts were removed from the data, allowing the removal of saccades (M = 132, SD = 61) and blinks (M = 369, SD = 139). Ocular artefacts were detected by the use of an algorithm that correlated ICA components to a template of a blink or a saccade and by correlations of the component activation time courses to the VEOG and HEOG observed difference waves.

Afterwards, a high-pass filter of 1 Hz and a low-pass filter of 60 Hz, using ERPLAB scripts and using a Butterworth filter (zero-phase), 6th order, appropriate to eliminate slow waves while keeping late transient auditory components, such as the N1 and P2, were applied to the EEG scalp signals. The EOG signals (HEOG and VEOG difference waves) were filtered with a 0.1 Hz high-pass filter and a 10 Hz low-pass filter. EEG was epoched from 500 ms before the first tone of the set to 2,500 ms after the start of set presentation, for each set. This meant epochs ended 700 ms after the last tone of each set for Load 5 condition, and 1,500 ms after the last tone of the set for Load 3 condition. This was made to allow direct comparison of the data for both load conditions. Doing this, the main concern would be to reject too many trials due to ocular or muscular activity after the end of the trials. However, counts of kept and rejected trial listed in Table 1 show that the number of rejected trials was not higher for the Load 3 condition. For the second set presentation, we only kept trials for which both sets were the same, such as not to include effects of tone deviation, or cognitive responses related to response selection that could be confounded with possible effects of expectation. No baseline correction was necessary because the high-pass filter eliminated slow drifts and centered the waveforms around 0 μV. As a further precaution, trials still containing large deviations on the recomputed (post ICA) HEOG and VEOG difference waves were removed when the HEOG signal varied by more than 35 μV over a 300-ms period, and/or if the VEOG signal fluctuated by more than 50 μV over a 150-ms period. Scalp-EEG electrodes with non-eye movement-related artifacts in the EEG signals (signal exceeding ±100 μV) were interpolated (at the exception of FCz, the electrode of interest), only if seven channels or fewer needed interpolation, otherwise the trial was excluded from further analysis.

Tableau 1 Mean, standard deviation, and range of the number of trials retained, and excluded (in parenthesis) for each condition across both experiments. Note there were about half as many trials in Set 2 as in Set 1 because we included only trials in which the second sequence was the same as the first.

	Mean	Standard deviation	Range
Load 3 - Set 1	196.37 (3.00)	3.91 (3.74)	181–200 (0–19)
Load 3 - Set 2	98.27 (1.30)	1.57 (1.29)	93–100 (0–4)
Load 5 - Set 1	196.67 (2.90)	7.11 (6.26)	161–200 (0–34)
Load 5 - Set 2	98.83 (0.97)	1.49 (1.10)	93–100 (0–4)

The grand-average ERP waveforms can be seen in Figure 2.2. An activity can be seen before the onset of the first sequence. This was caused by a fixed delay of 500 ms between the apparition of the fixation cross and the presentation of the first tone. We quantified the amplitude of the N1 and P2 components by computing the mean amplitude in a window centered on the peak of each component in the grand average waves at FCz for each participant. For each component, the measurement window was set to be approximately the width of the component measured between the time points at which the component was half of its maximum amplitude (so-called full width at half maximum, FWHM). The window was 75 ms wide for the N1 and 100 ms wide for the P2. The windows were thus 75–150 ms, 500–575 ms, 900–975 ms, 1305–1380 ms, and 1715–1790 ms relative to the start of each set presentation for the N1 component. For the P2 component, they were 180–280 ms, 600–700 ms, 1010–1110 ms, 1420–1520 ms, and 1825–1925 ms, relative to the start of each set presentation. Theses latencies were chosen by looking for the peak around 100 ms after tone presentation for the N1 and 160 ms for the P2, and by adjusting the window to be centered on the ERP. Because the first tone presented (no matter in which phase) proved to have a specific electrophysiological signature that was equivalent in all sequences and hence provided no differential information, we did not include it in our analyses.

We focused on the auditory components for the Load 5 condition, because it had more tones to compare, but checked if the pattern was consistent for Load 3. The mean amplitudes for each participant for the tone positions 2, 3, 4, and 5, were submitted to ANOVAs that included a specific contrast designed to test the hypothesis that the second and fourth tones would be different from the third and fifth. This pattern was encoded in the contrast weights (+1 -1 +1 -1) over tone positions.

We performed an ANOVA on the amplitude of each ERP component of interest (the N1 and the P2), testing for an interaction between tone position and set presentation. If such an interaction exists, it means that the pattern of amplitudes across the tones differs between set presentations. That is, we expected to observe an amplitude modulation when tones carried information about the sequence length, in the mixed experiment, but not when sequence length was fixed, in the blocked experiment. We also reported the main effects of set presentation and of tone position. An ANOVA with a planned contrast (comparing the second and fourth tones to the third and fifth, contrast of +1 -1 +1 -1) was performed to verify the hypothesis that the differences in amplitude patterns arose from the second and fourth tone being informative about sequence length. We checked the contrast for each set presentation of each experiment. Significant results indicate that the average amplitude of the component for the second and fourth tones was different than the average for the third and fifth tones. Because tone repetition[2] led to a decrease in amplitude, both for the N1 and the P2, during the presentation of the set, it could lead to a significant contrast because the third and fifth tones were later in the sequence (more repetition attenuation) than the second and fourth tones. The contrast comparison between sets overcomes this difficulty because they presumably included equivalent tone repetition attenuation. Repetition effects being the same between the two sets, if the information carried by a tone increases N1 and/or P2 amplitude, we should a difference in the contrast between the two set presentations only in the mixed experiment.

We also performed analyses that compared the amplitudes for the 3rd and 4th tones. These analyses allowed us to support the results from the ANOVA with the planned contrast. The fourth tone should elicit ERP components more attenuated by repetition than the third tone. However, an increase in N1 and/or P2 for the fourth tone provides evidence for the informativeness of that tone. We should see this only in the first set presentation of the mixed experiment, when the presentation of the fourth tone indicates that a fifth tone will be presented.

---

[2] Tones in a sequence all had different frequencies, hence there was no repetition. However, tones with closer frequencies might show attenuation effects similar to exact tone-repetition effects.

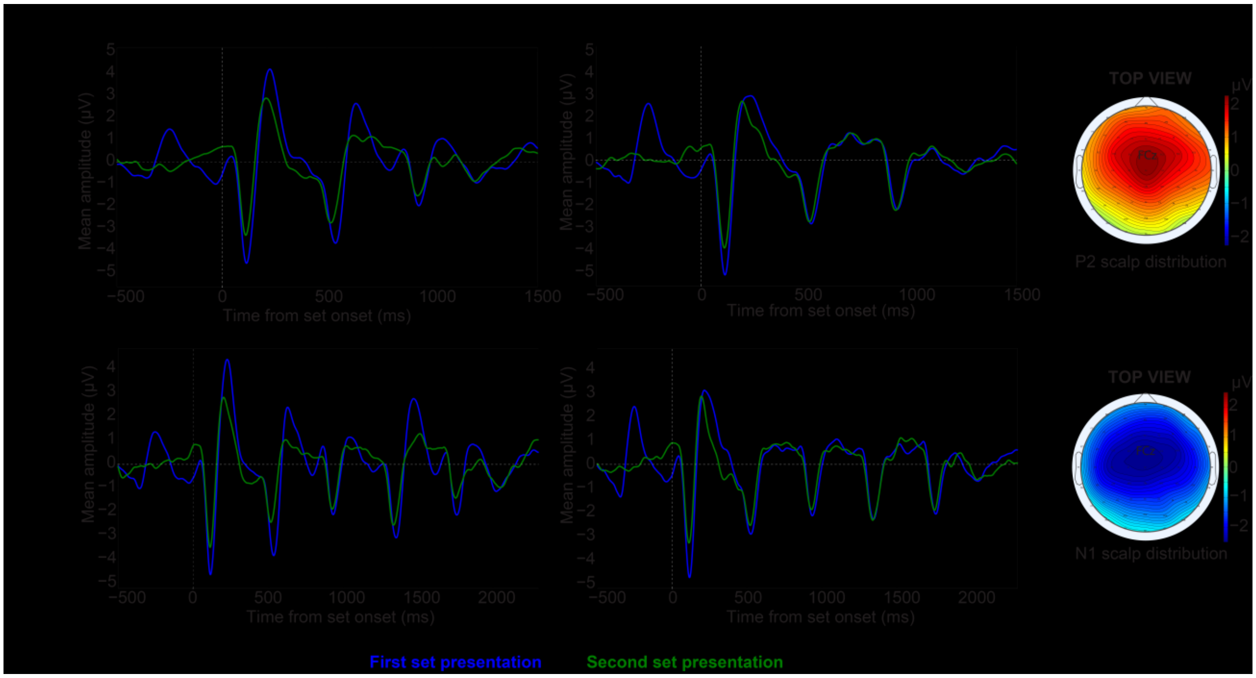

Figure 2.2 ERP waveforms at FCz for Load 3 and Load 5 condition, for both experiments: mixed experiment (Experiment 1) and blocked experiment (Experiment 2), and for both set presentation. ERPs were time-locked to the start of each set presentation. Scalp distributions for typical N1 and P2 in this experiment are shown on the right of the waveforms by computing the mean amplitude in a 75 ms window around the N1 peak, or a 100 ms window around the P2 peak for the first tone presented. Different abscissa scales for the two load conditions were used to maximize waveform detail. Ordinate axes were cut 500 ms after the end of each set presentation for clarity purposes.

2.4 Results

2.4.1 Behavioral

We analysed behavioral data with repeated measure ANOVAs with Greenhouse-Geiser correction when sphericity assumptions were violated, and Bonferroni correction for paired comparisons. RT outliers were removed using the Van Selst and Jolicœur (1994) method.

For Experiment 1 (mixed loads), a one-way repeated-measures ANOVA factoring load (3 levels: Load 1, Load 3, and Load 5) showed significant differences in accuracy between loads, $F(2, 28) = 22.40$, $p < .001$, $\eta_p^2 = .61$, corrected using Greenhouse-Geisser ($\varepsilon = 0.61$). Post-hoc comparisons with Bonferroni correction showed accuracy for Load 5 ($M = 89.7\%$, $SE = 2.1\%$) was significantly lower than for Load 1 ($M = 97.3\%$, $SE = 1.0\%$) and Load 3 ($M = 97.2\%$, $SE = 0.9\%$), $p < .001$ in all cases. RTs were not analyzed by load, nor condition, because of the delay imposed before accepting a response, but we report the overall mean as a rough indication of general response speed. Mean RT across loads was 442 ms ($SE = 35$), following the response cue, which occurred 500 ms after the end of the presentation of the second sequence.

Similar to Experiment 1, a one-way repeated-measures ANOVA for Experiment 2 showed significant differences in accuracy between loads, $F(2, 28) = 30.48$, $p < .0001$, $\eta_p^2 = .68$, corrected using Greenhouse-Geisser ($\varepsilon = 0.61$). Average accuracy was 96.7% ($SE = 1.0\%$) for Load 1, 96.1% ($SE = 0.8\%$) for Load 3, and 86.9% ($SE = 1.7\%$) for Load 5. A post-hoc comparison with Bonferroni correction showed accuracy for Load 5 was significantly smaller than for Load 1 and Load 3, $p < .001$ in all cases. Mean RT across load conditions was 370 ms ($SE = 20$), computed from the presentation of the response cue.

2.4.2 EEG

2.4.2.1 P2 component

Mean amplitudes of the P2 component for Load 5 are shown in Figure 2.3 for each experiment, combination of tone position, and set presentation. ANOVAs were corrected with Greenhouse-Geisser correction when sphericity was violated, as determined by Mauchly's test of sphericity (Mauchly, 1940). We expected the second and fourth tone of the Load 5 condition, and the second tone of Load 3 to show larger P2 amplitude in the first set presentation of the mixed experiment, but not in any other set presentation.

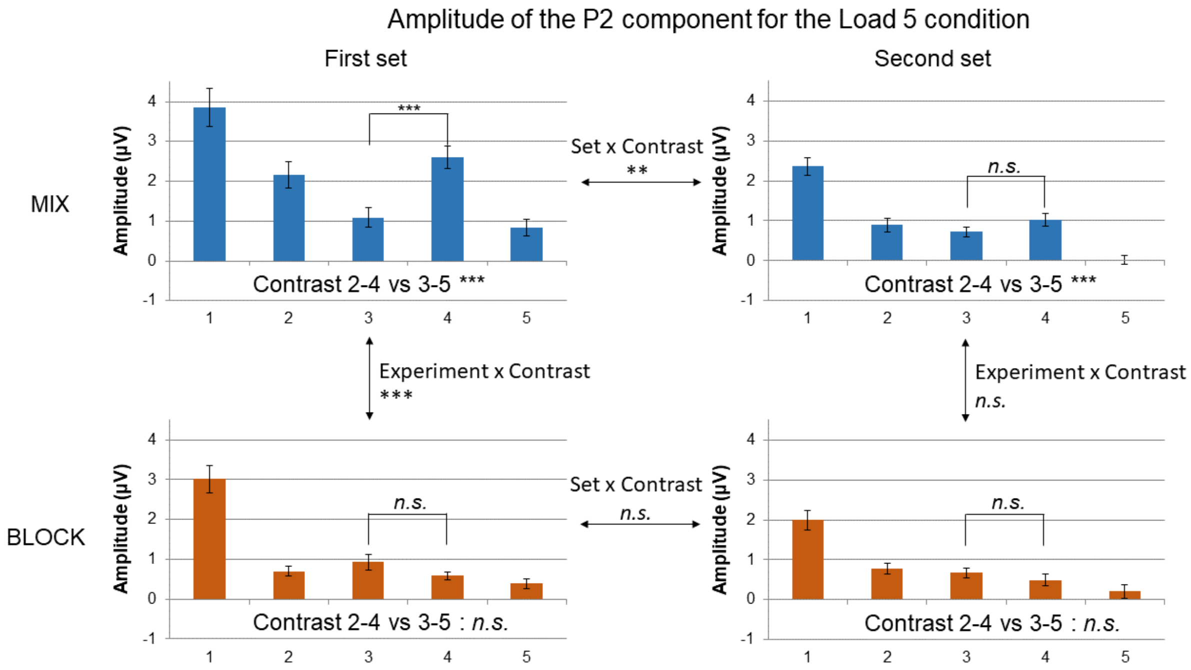

Figure 2.3 Mean P2 amplitude at electrode FCz for each tone position, each set, and each experiment for the Load 5 condition. Error bars show the standard error. Contrast 2-4 vs 3-5 refers to the contrast 0 1 -1 1 -1, applied to Tone positions 1 to 5. Differences in amplitude between the third and fourth tones, which are the most relevant comparisons, are illustrated. Significant differences are marked * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. Non-significant differences are marked n.s. Significant contrast in the second set presentation is driven by the significant difference between P2 amplitude of the fifth tone and mean P2 amplitude for the second, third, and fourth tone, $p < .001$, see text.

For the mixed design, mean P2 amplitude for the last four tones was larger for Set 1 than for Set 2, $F(1, 14) = 28.01$, $p < .001$, $\eta_p^2 = 0.67$, and was significantly different between tones, $F(3, 42) = 20.08$, $p < .0001$, $\eta_p^2 = .59$. There was a significant interaction between Set and Tone position, $F(3, 42) = 4.86$, $p < .006$, $\eta_p^2 = .26$, meaning the pattern of P2 amplitudes over the four tones was different between the two sets. An ANOVA with a planned contrast (comparing the second and fourth tones to the third and fifth, contrast of +1 -1 +1 -1) showed larger P2 amplitudes for the second and fourth tone compared to the third and fifth tones, for Set 1, $F(1, 14) = 29.02$, $p < .0001$, $\eta_p^2 = .67$, as well as for Set 2, $F(1, 14) = 20.79$, $p < .001$, $\eta_p^2 = .60$. However, there was also a difference in the amplitude of the contrasts across the first and second tone sets, $F(1, 14) = 10.20$, $p < .007$, $\eta_p^2 = .42$, confirming what can be seen in Figure 2.3, namely that the high-low-high-low pattern was larger in the first set than in the second set.

For the blocked design, P2 amplitude was not different, overall, between sets, $F(1, 14) < 1$, $p = .38$, but was significantly different between tones, $F(3, 42) = 4.64$, $p < .007$, $\eta_p^2 = .25$. There was no significant interaction between Set and Tone position, $F(3, 42) < 1$, $p = .50$. Because of the differences in P2 amplitude between the tones, we tested the planned contrast. Planned contrast was not significant in Set 1, $F(1, 14) < 1$, $p = .90$, nor Set 2, $F(1, 14) = 1.45$, $p = .25$. There was no interaction of Set and planned contrast, $F(1, 14) = 1.35$, $p = .26$.

We also compared the contrast for each set presentation (Set 1 and Set 2) between experiments. Because the general decrease in amplitude within each set presentation should be similar across experiments, this comparison would highlight differences due to the experimental manipulations. For Set 1, the contrast was significantly different between experiments, $F(1, 28) = 23.02$, $p < .0001$, $\eta_p^2 = .45$, reflecting a significant contrast for the mixed experiment and a non-significant contrast for the blocked experiment described above. Contrasts for Set 2 were not significantly different between experiments, $F(1, 28) = 3.61$, $p = .07$.

We expected a significant contrast for the first set of the mixed experiment but not for any other set. However, the contrast was significant for the second set of the mixed experiment. Visual inspection of the data showed a large reduction in P2 amplitude for the fifth tone of the second set, and some suggestions in the results for a trend for decreasing amplitudes, overall, from the beginning to the end of the sequences. We decided to focus on the anticipated increase in amplitude from Tone 3 to Tone 4, which could not be explained by a general decrease in component amplitudes.[3]

Thus, we performed further analyses comparing P2 amplitude for tones at positions 3 and 4 to ensure effects were not due to a simple decreasing pattern, nor adaptation or refractoriness. A mixed ANOVA with Tone position (2 levels: Tone 3 and Tone 4) and Set (2 levels: Set 1 and Set 2) as within-subject factors, and Experiment (2 levels: mixed experiment and blocked experiment) as the between-subjects factor yielded a significant 3-way interaction, $F(1, 28) = 9.47$, $p < .005$, $\eta_p^2 = .25$, reflecting the increase in P2 amplitude from the 3rd to the 4th tone, only in the first set of the mixed experiment. Separate ANOVAs for each experiment revealed a significant interaction between tone position and set presentation for the mixed experiment, $F(1, 14) = 10.58$, $p < .006$, $\eta_p^2 = .43$, but not for the blocked experiment, $F(1, 14) < 1$, $p = .52$. The interaction for the mixed experiment was explained by a larger P2 amplitude for the fourth tone than the third in the first set presentation, $F(1, 14) = 37.42$, $p < .0001$, $\eta_p^2 = .38$, which was not present in the second set presentation, $F(1, 14) = 1.53$, $p = .24$. Main effects for the blocked experiment showed no difference in P2 amplitude by tone position, $F(1, 14) = 3.34$, $p = .08$, nor by set presentation, $F(1, 14) = 1.24$, $p = .28$.

The mean amplitude of the P2 for Load 3 condition, for each experiment, tone set, and each tone position is shown in Figure 2.4. These means were submitted to the same type of analysis as in previous sections, but with only 2 tone positions (Tone 2 and Tone 3). For the mixed experiment, P2 amplitude was globally larger in the first set presentation than the second, $F(1, 14) = 25.33$, $p < .0002$, $\eta_p^2 = .64$, and the amplitude was overall larger for the second tone than the third, $F(1, 14) = 27.88$, $p < .0002$, $\eta_p^2 = .66$. There was no Set × Tone position interaction in the mixed experiment, $F(1, 14) = 1.15$, $p = .30$, meaning that the P2 amplitude pattern was the same across set presentation. For the blocked experiment, there was no effect of Set on P2 amplitude, $F(1, 14) < 1$, $p = .93$, but there was an effect of Tone position, the second tone had a larger amplitude than the third, $F(1, 14) = 11.49$, $p < .005$, $\eta_p^2 = .45$. There was no Set × Tone position interaction, $F(1, 14) < 1$, $p = .41$.

---

[3] We thank the reviewers for this suggestion.

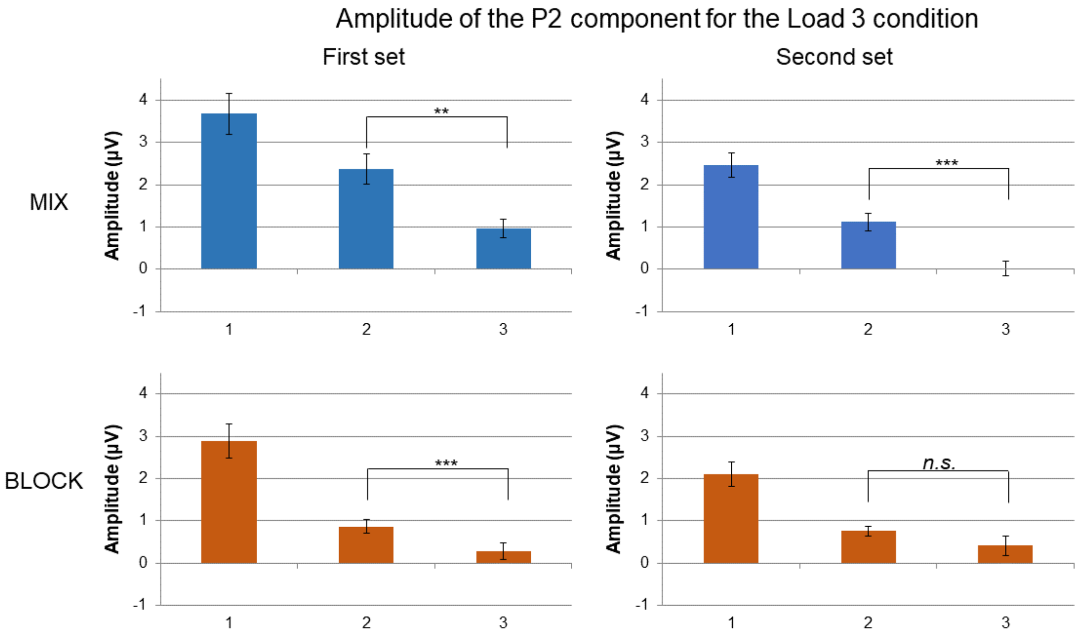

Figure 2.4 Mean P2 amplitude at electrode FCz, at Load 3, for each tone position, each set, and each experiment for the Load 3 condition. Error bars show the standard error. Significant differences are marked * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. Non-significant differences are marked *n.s.*

2.4.2.2 N1 component

Mean amplitudes of the N1 component for Load 5 are shown in Figure 2.5 for each experiment, combination of tone position, and tone sequence. ANOVAs were corrected with Greenhouse-Geisser correction when sphericity was violated. We expected, and tested for, the same patterns of results as in the analyses of the P2 component.

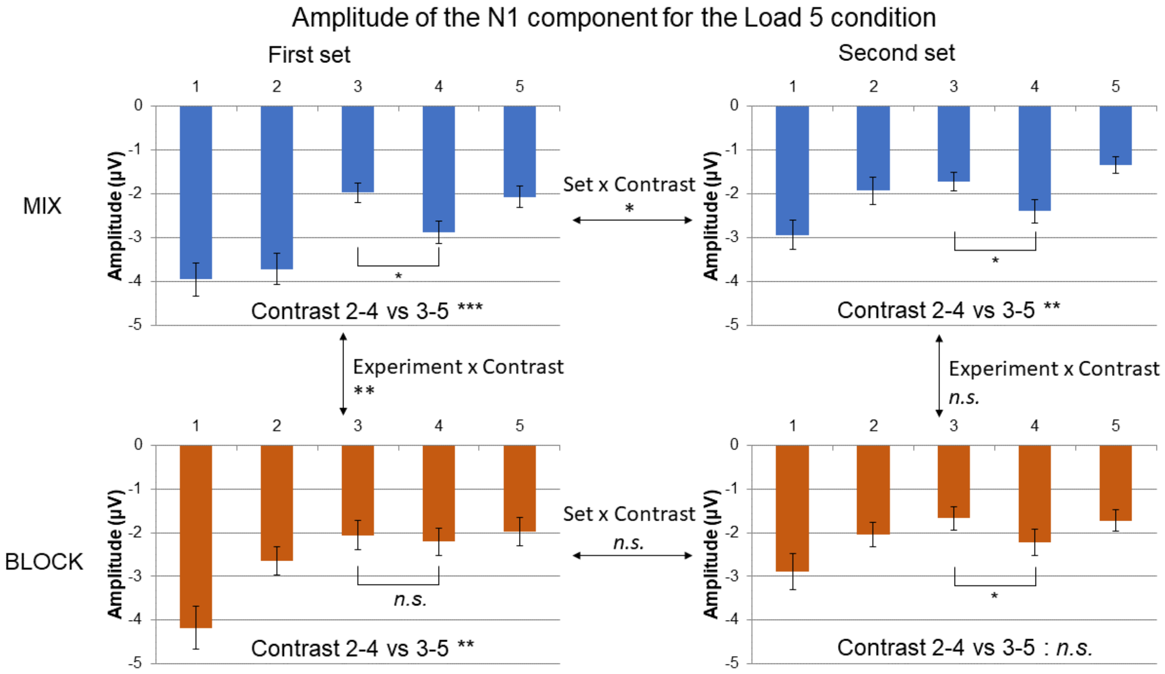

Figure 2.5 Mean N1 amplitude at FCz by tone position for each tone position, each set, and each experiment for the Load 5 condition. Error bars show the standard error. Contrast 2-4 vs 3-5 refers to the contrast 0 1 -1 1 -1, applied to Tone positions 1 to 5. Differences in amplitude between the third and fourth tones, which are the most relevant comparisons, are illustrated. Significant differences are marked * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. Non-significant differences are marked *n.s.*

For the mixed design, N1 amplitude was larger for the first set than the second set, $F(1, 14) = 20.37$, $p < .001$, $\eta_p^2 = .59$, and was significantly different between tones, $F(3, 42) = 14.37$, $p < .0001$, $\eta_p^2 = .51$. There was a significant interaction between Set and Tone position, $F(3, 42) = 6.55$, $p < .001$, $\eta_p^2 = .32$, meaning that N1 amplitude across tones was different between the two sets. An ANOVA with planned contrast (comparing the second and fourth tones to the third and fifth, contrast of +1 -1 +1 -1) showed larger N1 amplitudes for the second and fourth tone compared to the third and fifth tones, for the first sequence of tones, $F(1, 14) = 28.41$, $p < .0002$, $\eta_p^2 = .67$, as well as for the second sequence, $F(1, 14) = 8.94$, $p < .01$, $\eta_p^2 = .39$. There was an interaction between the contrasts, $F(1, 14) = 4.78$, $p < .05$, $\eta_p^2 = .25$, confirming what can be seen in Figure 2.5, namely that the high-low-high-low pattern was larger in the first set than in the second set.

For the blocked design, N1 amplitude was not different, overall, between sets, $F(1, 14) = 3.01$, $p = .10$, but was significantly different between tones, $F(3, 42) = 5.02$, $p < .005$, $\eta_p^2 = .26$. There was no significant interaction between Set and Tone position, $F(3, 42) = 1.16$, $p = .34$. Because of the differences in N1 amplitude between the tones, we tested the planned contrast. Planned contrast was significant in the first set, $F(1,14) = 12.67$, $p < .004$, $\eta_p^2 = .48$, but not in the second set, $F(1,14) = 3.61$, $p = .07$. There was no interaction between the contrasts of the two sets, $F(1,14) < 0.01$, $p = .95$.

As in the P2 analyses, we also compared the contrast of each set presentation between experiments. For the first set presentation, the contrast was significantly different between experiments, $F(1, 28) = 10.46$, $p < .004$, $\eta_p^2 = .27$, reflecting the significant contrast for the mixed experiment and the non-significant contrast for the blocked experiment described above. Contrasts for the second set presentation were not significantly different between experiments, $F(1, 28) < 1$, $p = .53$, as expected.

The larger contrast for the first set presentation of the mixed experiment partially supported our hypothesis. However, the contrast for the second set presentation of the mixed experiment was significant, which was not expected. Similar to the analysis conducted on the P2 component, we compared N1 amplitudes for tones at positions 3 and 4 to ensure that the effects were not due to the smaller N1 amplitude in response to the fifth tone, a simple decreasing pattern, nor adaptation or refractoriness. A mixed ANOVA with Tone position (2 levels: Tone 3 and Tone 4) and Set (2 levels: Set 1 and Set 2) as within-subject factors, and Experiment (2 levels: mixed experiment and blocked experiment) as the between-subjects factor yielded a non-significant 3-way interaction, $F(1, 28) = 1.90$, $p = .18$. For the mixed experiment, N1 amplitude was marginally larger in the first set than the second set, $F(1, 14) = 4.11$, $p = .06$, and was larger for the fourth tone than the third tone, $F(1, 14) = 9.75$, $p < .01$, $\eta_p^2 = .41$. There was no Set × Tone interaction, $F(1, 14) < 1$, $p = .43$, meaning that the N1 for the fourth tone was larger than for the third tone similarly in both sets. For the blocked experiment, there was no effect of set presentation on N1 amplitudes, $F(1, 14) < 1$, $p = .40$, but the fourth tone had a larger N1 amplitude than the third tone overall, $F(1, 14) = 6.96$, $p < .02$, $\eta_p^2 = .33$. There was no Set × Tone interaction, $F(1, 14) = 1.24$, $p = .28$, meaning that the N1 for the fourth tone was larger than the N1 for the third tone by the same amount in both sets.

The mean amplitude of the N1 for Load 3 condition, for each experiment, tone set, and each tone position is shown in Figure 2.6. The subject means were submitted to the same type of analysis as in previous sections, but with only 2 tone positions (Tone 2 and Tone 3). For the mixed experiment, N1 amplitude was globally larger in the first presentation than the second, $F(1, 14) = 19.32$, $p < .001$, $\eta_p^2 = .58$, and the amplitude was overall larger for the second tone than the third, $F(1, 14) = 40.97$, $p < .0001$, $\eta_p^2 = .74$. There was no Set × Tone position interaction in the mixed experiment, $F(1, 14) = 3.79$, $p = .07$, meaning that the N1 amplitude pattern did not differ significantly from the first to the second set. For the blocked experiment, there was not effect of set presentation on N1 amplitude, $F(1, 14) = 1.52$, $p = .24$, as well as no overall effect of tone position, $F(1, 14) = 2.57$, $p = .13$. There was a Set × Tone position interaction, $F(1, 14) = 5.21$, $p < .04$. This interaction was caused by the second tone having a larger N1 amplitude than the third for the first set presentation, $F(1, 14) = 11.95$, $p < .004$, $\eta_p^2 = .46$, but not for the second set presentation, $F(1, 14) < 1$, $p = .96$.

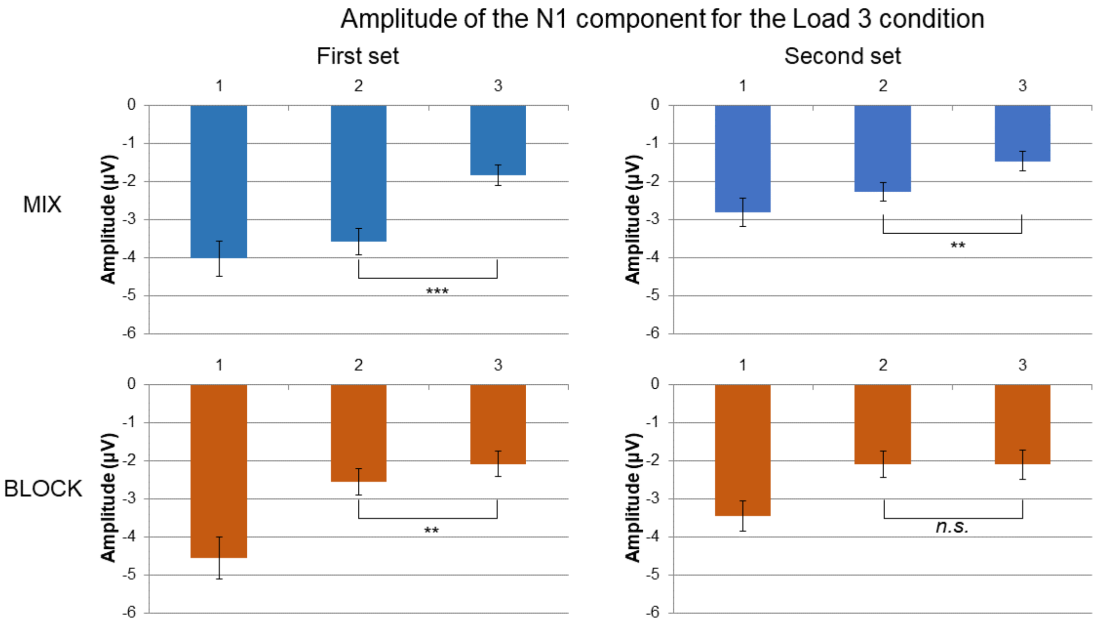

Figure 2.6. Mean N1 amplitude at FCz by tone position for each tone position, each set, and each experiment for the Load 3 condition. Error bars show the standard error. Contrast 2-4 vs 3-5 refers to the contrast 0 1 -1 1 -1, applied to Tone positions 1 to 5. Significant differences are marked * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. Non-significant differences are marked *n.s.*

2.4.3 Independent Component Analysis (ICA) of the ERP results

One may wonder whether the amplitude modulations we found in the P2 component might actually reflect modulations of an early P3a[4]. Our P2 component had a peak 230–275 ms following tone onset, which is 50–95 ms later than a typical P2, with a peak at about 180 ms. We note that the peak of the N1 component was also delayed and occurred 112–155 ms following tone onset, which was 12–55 ms later than the standard 100 ms latency. These delays may be caused, at least in part, by properties of the stimuli, such as the on-ramp time, or delay imposed by technical constraints such as the delay in onset of the sound at transducer, and possible propagation delay in the pneumatic coupling between the transducer and the ear.

It is, however, possible to account for the presence of a P3a component in our data if we supposed it could be elicited by task switching (Barcelo et al., 2006). The second and fourth tone could indicate that, instead of maintaining the memory, the memory should be extended, which could be similar to a task switch. Studies have also shown that P3a amplitude could be larger for stimuli that are deemed more significant due to information available from top-down processes, such as familiarity (Bobes et al., 2007), or bias in the subjective probability representation in oddball tasks (Horváth et al., 2008). If our positive-going wave was a P3a, it could reflect the contextual significance of the stimuli because the second and fourth tones do not become informative due to bottom-up processes, but due to top-down processes, including the knowledge of possible sequence lengths retrieved from long-term memory. This P2/P3a dilemma is described by Neszmélyi and Horvath (2017). In their study, the auditory N1 was followed by a positivity (P2/P3a), peaking at 236 ms in a passive listening condition. However, the positivity peaked at around 180 ms when sounds were self-generated, suggesting that the ERP component observed in the passive condition might reflect a P3a caused by attentional processes following a temporally unpredictable sound.

In order to address this issue, we performed an ICA of the ERP data, which can often separate spatially and temporally overlapping patterns of brain activity. For example, some studies used it to separate the P3a from the P3b in attentional blink experiments (Dell'Acqua et al., 2015, 2016). A recent study in our laboratory used this method to separate cognitive processes during auditory short-term memory retrieval (Simal & Jolicoeur, 2020).

---

[4] We thank the reviewers for this suggestion.

The data used for the decomposition was a concatenated matrix of 64 rows (each row corresponding to one channel), and a number of columns corresponding to the number of participants (N = 30) × the number of conditions (4 conditions used: Load 3 – First set, Load 3 – Second set, Load 5 – First set, Load 5 – Second set) × the number of measurements in a trial (1,536 time points; 3,000 ms with a sampling rate of 512 Hz). Thus, the matrix had 184,320 time points (30 × 4 × 1,536). Although Load 3 condition had response-related processes in the analysis window, we chose to include it to improve the quality and representativeness of the decomposition. Because the decomposition was made on a data matrix containing all participants and conditions, we obtained a fixed set of scalp maps (one per component) that applied to all participants and conditions. We used custom MATLAB code, as well as EEGLAB functions (Delorme & Makeig, 2004).

The first step was to restrict the solution space. Contrary to the PCA method that aims to group as much as possible the activity into each component, ICA aims to separate it. In our experience, doing a full ICA decomposition (here, into 64 components) results in most components reflecting the EEG activity of one participant and/or one condition. It is important to restrict the solution enough as to avoid overextraction (Fava & Velicer, 1992), but not too much as to avoid underextraction, which would degrade solution quality (Fava & Velicer, 1996). To determine the optimal subspace, we performed a singular value decomposition (SVD) on our concatenated ERP data matrix, and used the scree test to choose the number of components to retain. The scree test (Cattell, 1966; D'agostino Sr & Russell, 2005) is a technique consisting of plotting either the eigenvalues, or the variance explained by each component, and to find the point at which the variance of the components levels off, which is, according to Cattell, when the components are mainly measuring random error. In Figure 2.7 we show the scree plot from the SVD decomposition, which suggests the ideal cut-off was 8 components.

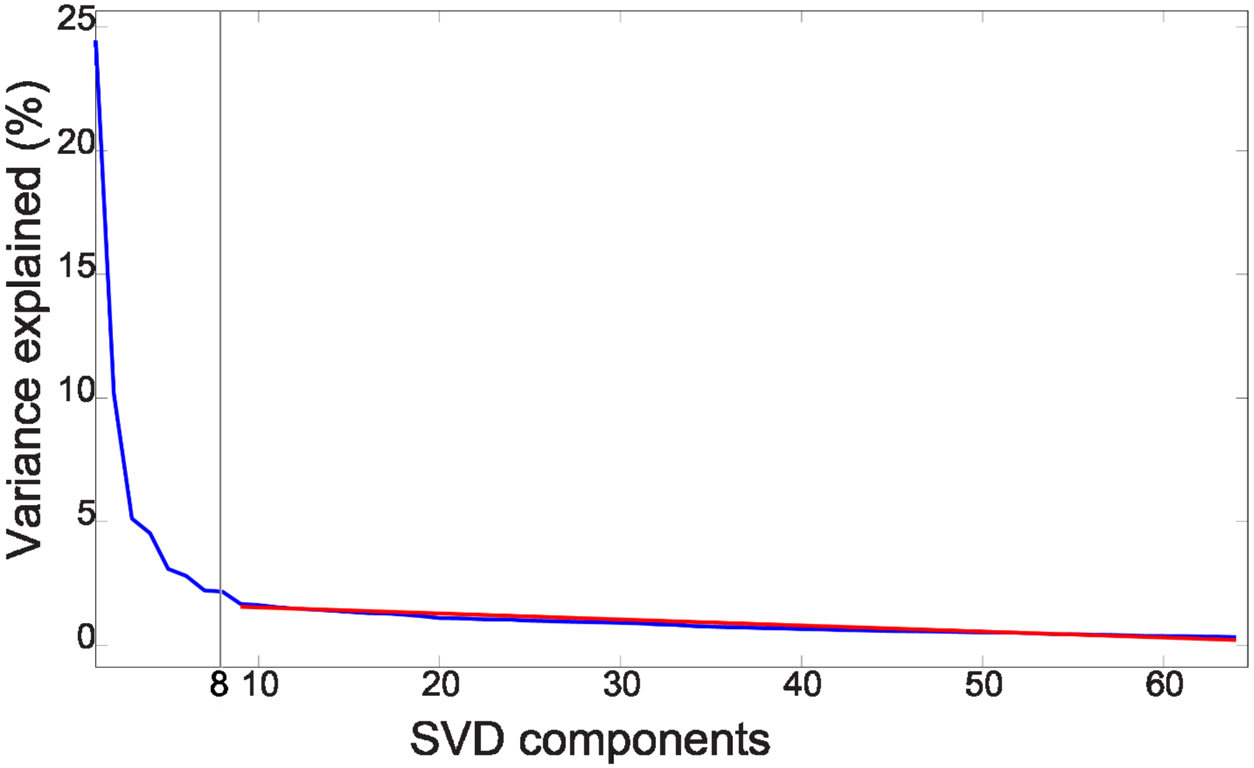

Figure 2.7. Scree plot of the variance explained by each singular value decomposition (SVD) component. The red line indicates the components for which variance levels off. The grey line at component number 8 indicates the best cut-off.

We performed an Infomax ICA decomposition (Bell & Sejnowski, 1995) after a dimensionality reduction to 8 components by the means of a PCA using the ‘pca’ flag in EEGLAB runica function. Because comparisons between tones were clearer for the Load 5 condition, we only present the ICA for Load 5. Figure 2.8 shows all the time course of the activations along with the scalp distribution of those activations.

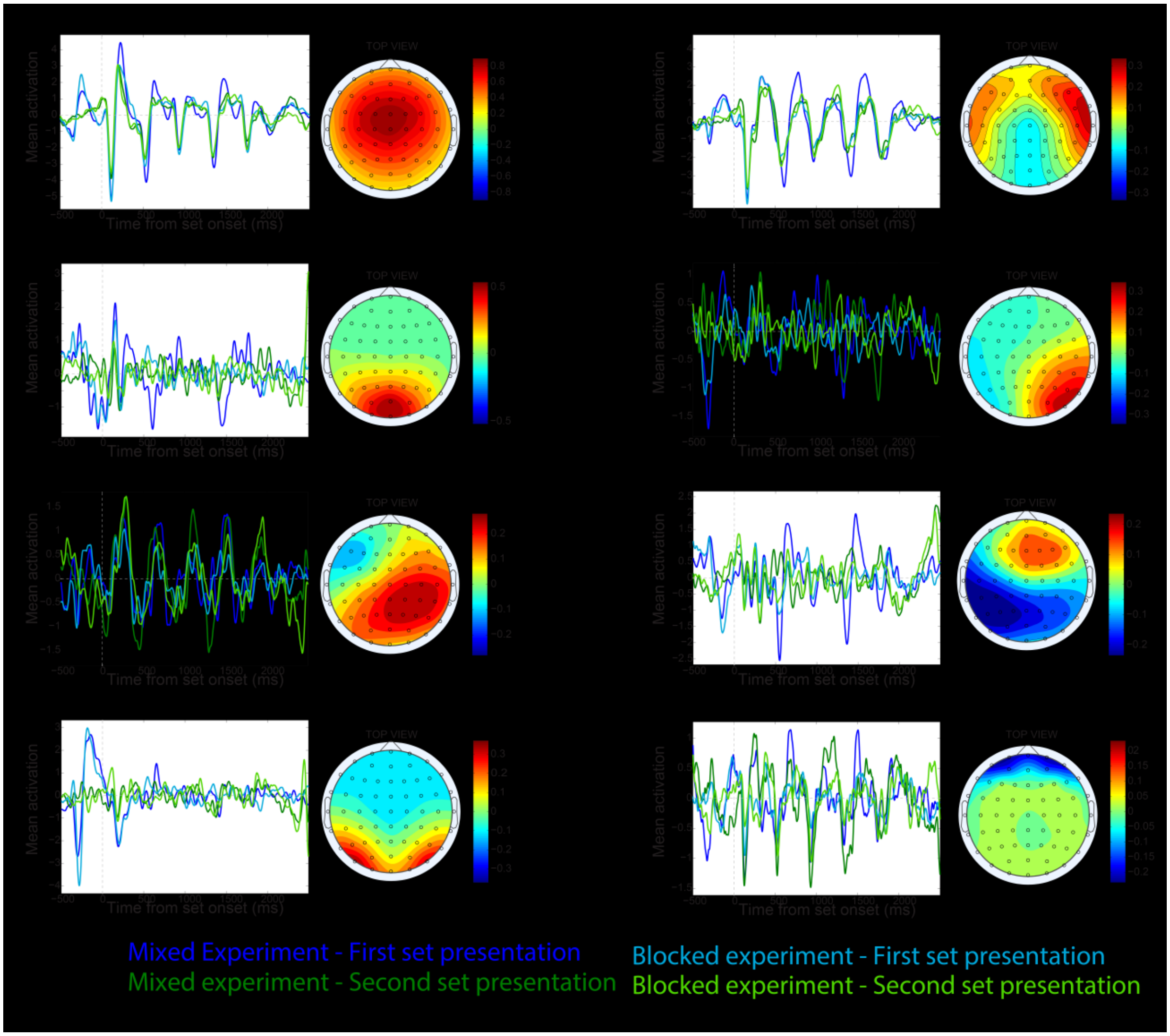

Figure 2.8. Independent components analysis of the ERP results. Shown here, for each component, is the fixed scalp distribution for that component and grand-average time courses for Load 5 for each condition (First set presentation, and Second set presentation), and Experiment. The original ERP data can be reconstructed as the product of the time courses and the associated scalp distribution, summed over all components.

The first ICA component, IC1, had a scalp distribution and a time course coherent with the main auditory ERP. Measurement using the same time windows as on the ERP showed significant activation for all N1 and P2 responses, $t(29) > 3.05$, $p < .005$, except for the P2 to the fifth tone of the second set presentation, $t(29) = 0.79$, $p = .44$. ANOVAs with contrast similar to those used on the ERPs showed a similar pattern. For the mixed experiment, activations were larger in the P2 time range of the second and fourth tones than of the third and fifth tone, $F(1, 14) = 31.72$, $p < .0001$, $\eta_p^2 = .69$, but there was no interaction of set and contrast, $F(1, 14) = 3.00$, $p = .10$. The activations in the time range of the second and fourth N1 time windows were also larger than those in the third and fifth N1 time windows, $F(1, 14) = 28.04$, $p < .001$, $\eta_p^2 = .67$, with an interaction between contrast and set $F(1, 14) = 6.14$, $p < .03$, $\eta_p^2 = .30$. This interaction arose from a larger effect in the contrast for the first set presentation, $F(1, 14) = 27.56$, $p < .001$, $\eta_p^2 = .66$, than for the second set presentation, $F(1, 14) = 7.35$, $p < .02$, $\eta_p^2 = .34$. For the blocked experiment, activations in the P2 time range of the second and fourth tones were not different than those in the P2 time range of the third and fifth tones, $F(1, 14) = 1.08$, $p = .32$. There was also no interaction of Set × Contrast, $F(1, 14) < 1$, $p = .82$. For activations in the N1 time ranges, the contrast was significant overall, $F(1, 14) = 7.78$, $p < .02$, $\eta_p^2 = .36$, and there was no Set × Contrast interaction, $F(1, 14) < 1$, $p = .83$. These results for IC1 indicated that this component included a large part of the auditory ERP, including some of the N1 and P2 modulations.

IC2 had a scalp distribution with activation at posterior electrodes with a maximum at electrodes POz and Oz. Comparison of the activation time course for IC2 with the ERP waveform for Load 5, first set presentation (see Figure 2.S3), showed the activations occurred mainly after the N1, for the first, second, and fourth tones of the first set presentation in the mixed experiment.

IC3 had significant activations in the time range of the P2 responses, $t(29) > 2.18$, $p < .04$, except for the P2 for the fifth tone of the first set presentation, $t(29) = 0.87$, $p = .39$. It was also significantly different from zero for the third N1 of the second set presentation, $t(29) = 2.30$, $p < .03$, and for the fourth N1 of the second set presentation, $t(29) = 4.34$, $p < .0002$. This IC did not reflect N1 or P2 modulations. For the mixed experiment, contrast on the activations in the P2 time ranges were not significant overall, $F(1, 14) = 2.56$, $p = .13$, and there was no Set × Contrast interaction, $F(1, 14) = 2.96$, $p = .11$. In the N1 time ranges, there was also no overall effect of the contrast, $F(1, 14) < 1$, $p = .68$, and no Set × Contrast interaction, $F(1, 14) < 1$, $p = .90$. For the blocked experiment, the contrast on the activations in the P2 time ranges were not significant overall, $F(1, 14) < 1$, $p = .79$, and there was no Set × Contrast interaction, $F(1, 14) = 1.46, p = .25$. In the N1 time ranges, there was also no overall effect of the contrast, $F(1, 14) < 1, p = .75$, and no Set × Contrast interaction, $F(1, 14) < 1$, $p = .34$. In summary, this component seems to represent a subcomponent of the P2, slightly lateralized to the right, that is not sensitive to the informativeness of the tones.

The fourth component, IC4, isolated the visual ERP occurring before the first set presentation in both experiments. The scalp distribution has maximal activation at electrodes PO7 and PO8 and main activations before the onset of the first tone of the first set presentation.

The fifth component, IC5, seemed to reflect the auditory ERP at electrodes T7 and T8. However, the superposition of the ERP waveform and the activation for this IC shows that latencies for each component are longer in IC5 (see Figure 2.S4). The activations corresponding to the N1 were delayed by approximately 50 ms, and activations corresponding to the P2 were delayed by approximately 115 ms. We used these delays to determine the new measurement windows. Because the activations in this IC seemed to be modulated the same way the N1 and P2 were modulated in the ERPs, we used the same contrast analysis, comparing the activations in response to the second and fourth tone to those in response to the third and fifth tone (contrast of +1 -1 +1 -1). In the mixed experiment, the contrast for the positive activations occurring right after the P2 was, overall, significant, $F(1, 14) = 26.59, p > .0002, \eta_p^2 = .65$. There was a Set × Contrast interaction, $F(1, 14) = 9.11, p < .01, \eta_p^2 = .39$, arising from a significant contrast for the first set presentation, $F(1, 14) = 31.41$, $p < .0001, \eta_p^2 = .69$, but not for the second set presentation, $F(1, 14) = 1.03, p = .33$. For the negative activations occurring right after the N1, the contrast was overall significant, $F(1, 14) = 6.80, p < .03$, $\eta_p^2 = .33$. There was a significant Set × Contrast interaction, $F(1, 14) = 7.36, p < .02, \eta_p^2 = .34$, arising because the contrast was significant for the first set presentation, $F(1, 14) = 14.07, p < .003, \eta_p^2 = .50$, but not for the second set presentation, $F(1, 14) < 1, p = .35$. For the blocked experiment, the contrast for the positive activations occurring right after the P2 was overall not significant, $F(1, 14) = 2.72, p = .12$, and there was no Set × Contrast interaction, $F(1, 14) < 1, p = .97$. The contrast for the negative activations occurring right after the N1 was also not significant overall, $F(1, 14) < 1, p = .67$, with no Set × Contrast interaction, $F(1, 14) < 1, p = .62$. This fifth component seems to be related to auditory processes and modulated by tone informativeness. Its scalp distribution being inconsistent with a P3a gives an indication that the modulations are likely not a P3a. We elaborate on the implications of this component in the discussion.

IC6 had a peak of activation similar to the visual ERP before the first set presentation. There was no significant activation during the N1 and P2 time windows, $t(29) < 2.01$, $p > .05$, except for the time window of the fifth N1 of the second set presentation, $t(29) = 4.29$, $p < .0002$. This component does not seem to reflect cognitive processes related to audition.

IC7 had a dipolar distribution with maximal activation at right frontal electrodes (Fz, F2, F4), and at temporal and parietal left electrodes (P1, P3, P5, P7, TP7, T7, CP3, CP5). This component had significant activations during the N1 time window for the second, third, fourth, and fifth tone of the first set presentation, $t(29) > 2.14$, $p < .05$, as well as the fifth tone of the second set presentation $t(29) = 2.35$, $p <.03$. Activations were also significant in the P2 time window for the second and fourth tone of the first set presentation, $t(29) > 4.01$, $p < .0004$. Because the activations seemed different between experiments, we redid separate *t*-test for each experiment. For the mixed experiment the activations in the N1 time window were significant for the second, third, fourth, and fifth tone of the first set presentation, $t(14) > 2.16$, $p < .05$. An ANOVA with planned contrast showed that N1-related activation was larger for the second and fourth tone than for the third and fifth tone, $F(1, 14) = 18.44$, $p < .001$, $\eta_p^2 = .57$. Activations in the P2 time window were significant for the second, fourth, and fifth tone of the first set presentation, $t(14) > 2.44$, $p < .03$. There was no significant activation for tones of the second set presentation. For the blocked experiment, activations in the N1 time window were significant for the second tone of the first and the second set presentation, $t(14) > 2.25$, $p < .05$. Activation was also significant in the P2 time window of the fifth tone of the first set presentation, $t(14) = 3.48$, $p < .005$. Because most activations occur in the first set presentation of the mixed experiment and more prominently for the second and fourth tone, we suggest that this component reflects part of the N1 and P2 modulation to tone informativeness.

IC8 only had significant activations during the N1 time window for the first and third tone of the first set presentation, $t(29) > 2.05$, $p < .05$, and for the second, third, and fourth tone of the second set presentation, $t(29) > 2.48$, $p < .02$. This component seems to reflect part of the N1 activity that does not include modulations from tone informativeness.

2.5 Discussion

The results provide electrophysiological evidence that cortical responses vary depending on the degree to which a tone in a given tone position provided information concerning the length of the sequence within which it was presented. We suggested that the second and fourth tones in the Load 5 condition, and the second tone in the Load 3 condition were more informative than the third or fifth tones, because these tones carried more information about subsequent tones to be heard in the first presentation of the mixed experiment but not in any other presentations (second sequence in the mixed experiment, and both sequences in the blocked experiment). The results provided significant support for these hypotheses, particularly the patterns of P2 amplitudes for Load 5, which we discuss in a subsequent section.

In summary, for the Load 5 condition, which had more reliable results than for the Load 3 condition, we found the P2 in response to the second and fourth tones, deemed more informative, was larger than for the third and fifth tones in both sets of the mixed experiment. However, this difference was more pronounced in the first set presentation (see Figure 2.3), when sequence length was not known. For the blocked experiment, however, the amplitude increases for Tones 2 and 4 were attenuated or absent. Comparisons across experiments also confirmed that the increase of P2 amplitude for the second and fourth tone was largest in the first set presentation of the mixed experiment.

For the N1 component (see Figure 2.5), we found similar patterns as for the P2 component, the most important prediction for the 2nd and 4th tones to have larger N1s than for the 3rd and 5th tones was significantly larger in the first set presentation of the mixed experiment than in the first set presentation of the blocked experiment. This pattern was also not different for the second set presentation between both experiments.

There were other factors that also affected the relative amplitudes of the N1 and P2 components across tone positions, tone sets, and experiments (mixed vs. blocked), which we consider first. There was a systematic tendency for the amplitude of the N1 and P2 to be smaller for the last tone of the second presentation. We had not anticipated this effect, but it can be explained by the same general account as the anticipated increase in amplitudes for informative tones. Recall that tone sequences that were different in the second sequence compared with the first were created by transposing the positions of two adjacent tones. Thus, participants could know whether the presentations differed or not before that last tone. Either the tone before last was the same as in the first presentation, and the two sets would be the same, or it was different (corresponding to the tone presented last in the first sequence), and the last tone would be different as well. Consciously or unconsciously, participants did not need to pay attention to that last tone, and the reduction in the N1 and P2 amplitudes suggest that less attention was deployed to the last tone. This also means that results for the Load 3 condition are not as easy to interpret as for Load 5, because we find ourselves in a situation where two of the three tones were likely modulated by factors other than the informativeness of the tones for sequence length. An experiment with the presentation of four or six tones could be useful to avoid this interpretation difficulty in the Load 3 condition, as well as to allow the generalization of our results to other sequence lengths[5]. Another approach would be to replace one of the tones in the second sequence by a tone that was not in the first sequence in trials in which the first and second sequences are not the same, which would make the last tone as diagnostic as any other tone.

The results for P2 amplitude in Load 5 condition were clear. We observed larger amplitudes for the second and fourth tones during the first presentation in the mixed experiment, as expected if these tones carried information about more upcoming tones. This was not the case in the blocked experiment, because there was no new information provided by these tones given that sequence length was always known in advance (and hence none of the tone positions added new information). Although the contrast (+1, -1, +1, -1; for Tones 2 to 5) testing the main hypothesis was significant for the second presentation, it is clear this was caused by the small amplitude of the last tone, which we discussed earlier. Moreover, there was no difference in P2 amplitude nor contrast in the second presentation between experiments. Comparing the P2 amplitudes for the third and fourth tones showed the expected results (the third, less informative tone had a smaller amplitude than the fourth, signaling the forthcoming fifth tone), which is not compromised by the small amplitude for the last tone. For Load 3, comparisons were made only between the second and the third tones, which had systematically smaller N1 and P2 amplitudes in the second set, as discussed above. This caused the results to be less clear and more difficult to interpret. The general pattern, however, provided some support for our hypothesis: P2 amplitude for the second tone in the first sequence of the mixed experiment was larger than P2 amplitude for the second tone in any other presentation. We suggest

---

[5] Examination of earlier unpublished experiments in our lab, informally verified that the expected patterns generalized to other sequence lengths. It would be useful to show this formally in future research.

this reflected the addition of information about sequence length, which occurs mainly in the first sequence of the mixed-loads experiment.

For the N1 component, the most important findings were that the second and fourth tones in the Load 5 condition were larger than for the third and fifth tones in the first sequence of the mixed experiment, and that this pattern was stronger than in the second sequence. Somewhat surprisingly, the amplitude pattern found for the N1 amplitudes showed larger N1 amplitudes for the second and fourth tones in the Load 5 condition in all presentations, although to a lesser extent the conditions other than the first sequence of the mixed experiment. However, the comparison between the N1 amplitude for the third tone and the N1 amplitude for the fourth tone, which eliminates the possibility that the effects are due to a simple decreasing pattern, adaptation or refractoriness, showed that there was no effect in the blocked experiment, as expected. In the mixed experiment, the fourth tone elicited a larger N1 than the third in both sets. In the Load 3 condition, N1 amplitude was larger for the second tone than for the third in all presentations, but larger overall in the first presentation of the mixed experiment. The effect of tone position on N1 amplitude in the second set of the mixed experiment is difficult to explain. Because the N1 is sensitive to many other processes than predictability, such as attention, repetition, tone intensity, and priming (see Näätänen & Picton, 1987, for a review), there might be some cognitive activity in this particular set presentation affecting the N1, which we did not expect. However, the differences in N1 patterns between the two experiments generally supported our hypothesis.

We have presented results from two closely related studies of auditory short-term memory that provide evidence that the P2, and to some extent, the N1, components (part of a complex of late-latency auditory components) are sensitive to disambiguating information concerning the length of rapidly-encoded tone sequences. Specifically, the P2 amplitude showed a relative increase at Tone 2 of 3-tone sequences (indicating that there would be more than 1 tone) and at Tones 2 and 4 of 5-tone sequences (indicating that they would be longer than 1 or 3 tones). Crucially, this effect was stronger in the experiment that randomized sequence length from one trial to another, when sequence length was not known in advance, and not when subjects performed the task in predictable blocks comprised of a single sequence length (the only parameter differing in the second experiment). In the latter case, the aforementioned tones (Tone 2 and Tone 4) did not carry any novel information concerning the length of sequences. Also telling is that this pattern of relatively greater P2 amplitude was clearly stronger during the presentation of the first set of any given trial of the mixed experiment, that is, during encoding (Set 1) but not at test (Set 2). The length of the second sequence was always known, having just been heard 2 seconds earlier in the first sequence. Therefore, no new information about sequence length was provided by Tones 2 or 4 in this case.

We wondered whether these effects on the amplitude of P2 could be thought of as violations of expectation. However, second and fourth tones were more expected than unexpected with a conditional probability of .66 for the second tone, and of .5 for the fourth tone (because in trials where they were at least 2 tones, there was a 50% chance that the trial would be Load 5). Another way to look at the results might be that the third and fifth tones were completely predictable once a second tone, or a fourth tone, was presented. In this view, the amplitude of predicted tones would be smaller than for tones whose occurrence was not fully predictable. However, this is essentially a restatement of the notion that the second and fourth tones convey more information about the length of the sequence than the third or fifth tones.

Unlike most previous descriptions of ERP amplitude as inversely proportional to the predictability of stimulation (e.g., the P3, Donchin, 1981; or the MMN, see May & Tiitinen, 2010, for a review, critique, and theory), for which less-frequent events produce a larger response, the relative increase in P2 amplitude described in the present work occurs in response to an expected number of elements (1, 3, or 5 tones) in a sequence, a sequence in which the selection from among six possible tones is randomized at every trial. There is therefore no pattern or expectation emerging from the low-level features or sequential occurrence of tones themselves (e.g., the repetition of a standard tone). Instead, the effects appear to be caused by a relatively abstract regularity (3 possible lengths of sequence) that must be extracted from the number of concatenated tones. The frequent confound of potential neural habituation encountered in relevant literature is thereby doubly circumvented by a selective increase in the P2 amplitude despite a random selection of tone frequencies and, that, the effects occur in response to an expectation concerning the 'supra-stimulation' feature of the number of elements in a sequence. It is also worth mentioning that our index is not a 'difference wave' like the 'repetition positivity' (RP) or the mismatch negativity (MMN) but rather a direct modulation of the N1 and P2 amplitude. Furthermore, the presentation of exactly the same sequences produced different patterns of relative amplitudes within the tone sequences, depending on whether the observer already knew the sequence length (i.e., in the 2nd sequence of a memory experiment, or when sequence length is fixed, within a block of trials). Thus, the effects we describe cannot be explained by neuronal adaptation in sequential presentations (see May & Tiitinen, 2010).

Importantly, the amplitude modulations we found occurred within 100 ms of tone onset (for the N1), and within a little more than 200 ms (for the P2), which likely precludes a conscious response triggered by tone onset. It is possible that a higher attentional weight could have been assigned to the second and fourth tones, in advance of their presentation, but it seems unlikely this would be a conscious strategy designed to enhance task performance. We suppose the main task during the first sequence was to encode the tones—their pitch—, in sequence, in preparation for the upcoming test, a few seconds later. The second and fourth tones are not particularly special, compared with the others, at the time of test. A memory task in which subjects only needed to remember the number of tones (1, 3, or 5) would have been trivially easy. We suppose that knowing how many tones were in the sequence was encoded implicitly in the sequence of varying pitches.

The possibility that the larger P2 amplitude reflect an amodal P3a response was examined using ICA. We were able to decompose the ERP data into subcomponent with distinct spatial and temporal activations. Activity was separated in an IC maximal at the vertex, accounting for the main activity seen in the ERPs, as well as visual activity at the start of the trial. More importantly, we were able to find ICs accounting for subcomponents of the N1 and the P2 that were sensitive to tone informativeness with a different topography than the fronto-central activation that is common to both the P2 and the P3a. Particularly, IC5, which showed maximal activation over temporal electrodes T7 and T8 reflected the larger N1 and P2 in response to the second and fourth tones (see Figure 2.8). Studies have demonstrated that the N1 has a complex generator structure (Näätänen & Picton, 1987) with both midline and temporal subcomponents (Wolpaw & Penry, 1975; Woods, 1995). The midline N1 component, is often differentiated from its subcomponents by the different processes required in a given task. Mid-temporal sites, such as those portrayed in IC5 are where the subcomponents N1a and N1c are maximal. Because the negative activations of this component peak later than the typical N1 component (see Figure 2.S4 for comparison), they likely reflect the N1c subcomponent (peaking 130–170 ms after tone presentation; Woods, 1995). To our knowledge, no study has examined possible multiple generators of the auditory P2, this component being mostly interpreted as part of the N1-P2 complex. While we cannot demonstrate that the positive activity observed in IC5 is a P2, the activations observed are likely generated in the auditory cortex. Moreover, these activations are sound-related and are modulated by informativeness the same way as the N1 and P2 components. Because the P3a does not have a distribution over temporal electrodes, this IC suggests that the modulations observed in the ERPs do not reflect a form of P3a.

In closing, we suggest the present results reflects brain activity related to contextually-determined disambiguating information on a very short timescale, even though this does not alter conscious awareness or overt strategy in the performance of the task. It has long been proposed that the brain is active in generating models, predictions, and extrapolations concerning the continuous stimulation impinging upon it, as well as creating complex and longer-term expectations and projections that aid in guiding behaviour. Examples of this abound and there has been considerable recent progress in the investigation of related phenomena (for a review, see Bendixen et al., 2012). Despite these recent advances in knowledge, there remains much to discover concerning the underlying mechanisms and their neural indices. We have presented here what we believe is compelling evidence that the P2 and N1 auditory components, as recorded during the encoding phase of an auditory short-term memory task, may be such an index given their apparent sensitivity to the contextually-regular lengths of tone sequences.

2.6 Acknowledgements

We thank Pia Amping for programming.

The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs program, and the Canada Foundation for Innovation awarded to PJ.

2.7 Supplementary figures

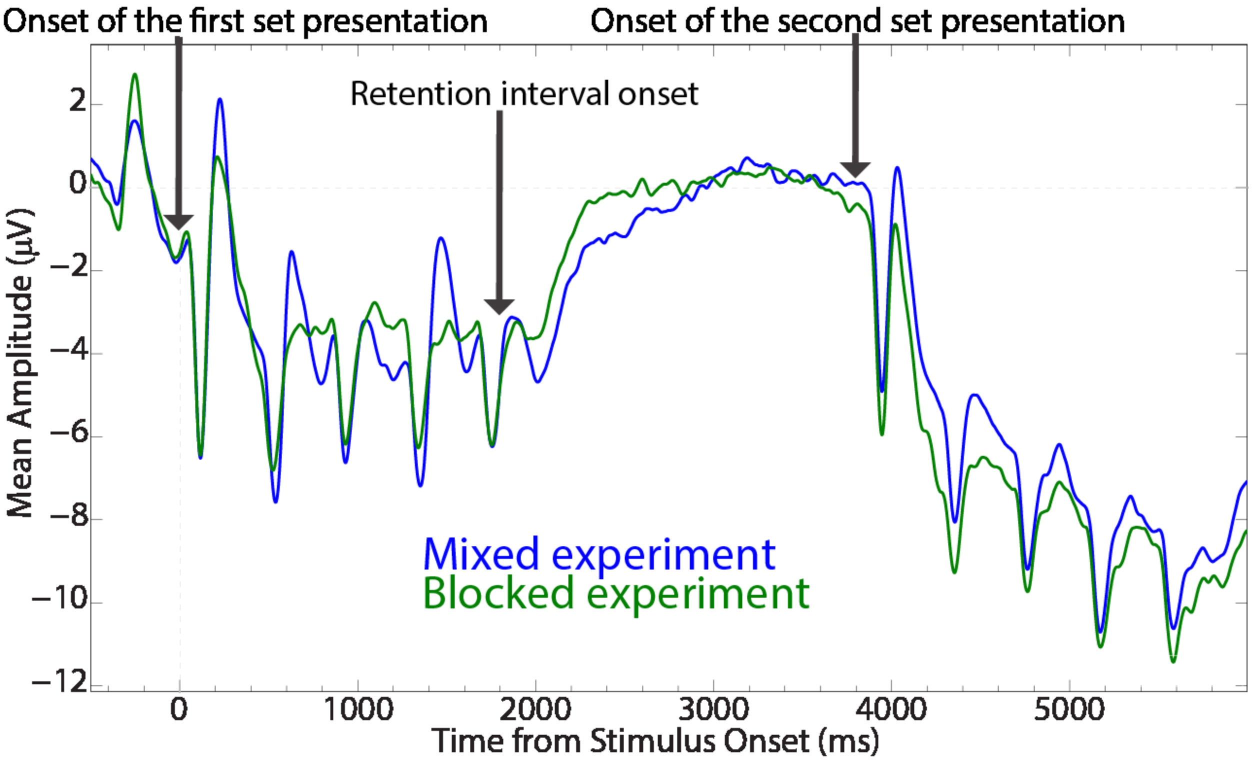

Figure 2.S1. ERP waveforms at electrode FCz, with a band pass filter 0.1–60 Hz instead of the 1–60Hz used for the analysis, for Load 5 condition, for both experiments: mixed experiment (Experiment 1) and blocked experiment (Experiment 2). ERPs are time-locked to the onset of the first set presentation. Please note that the blocked experiment waveform was based on 14 participants instead of 15 because of artefact rejection.

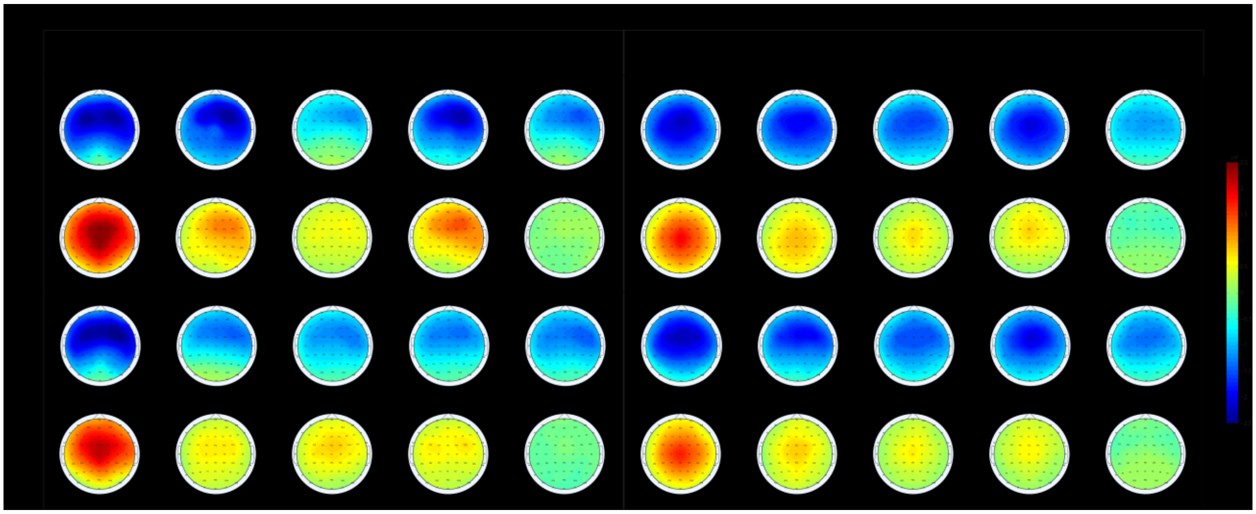

Figure 2.S2. Scalp distribution of the N1 and P2 for each tone in each set and each experiment.

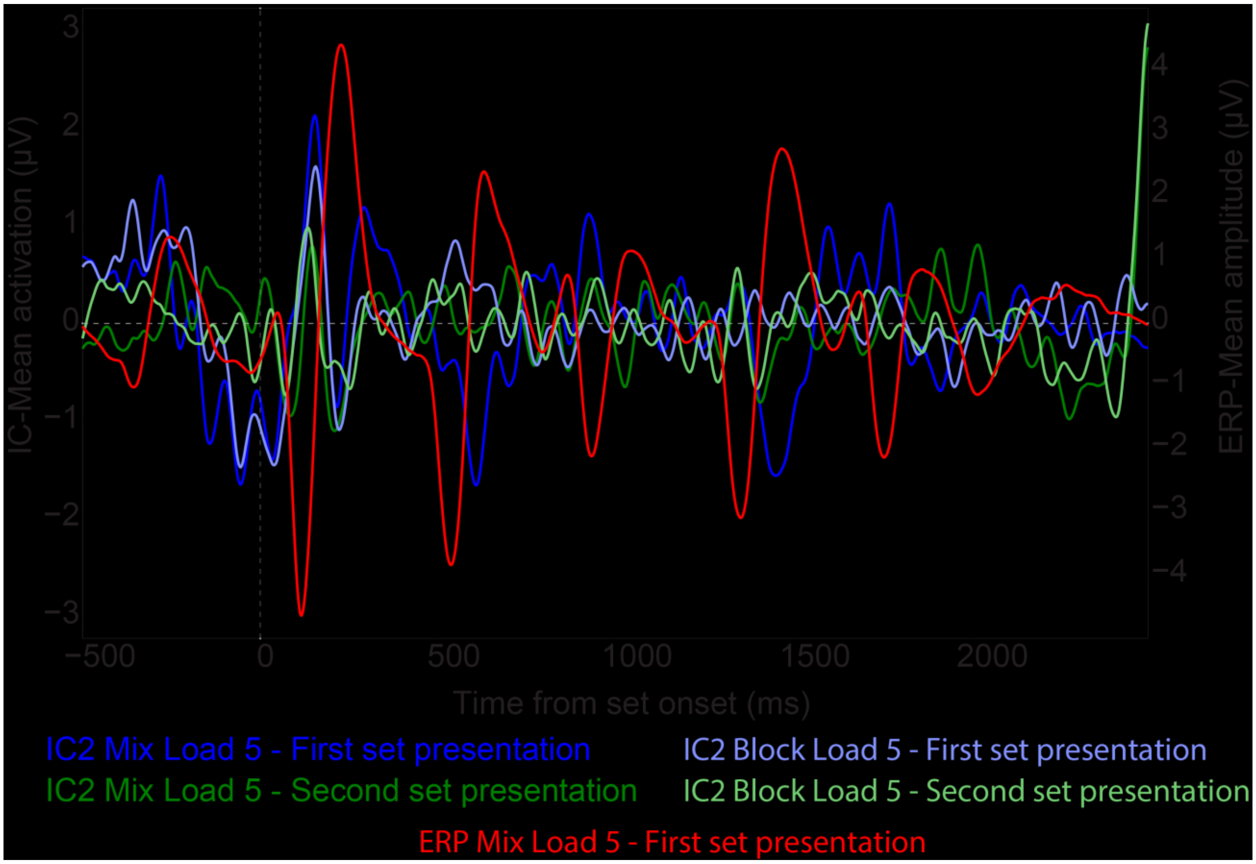

Figure 2.S3. Comparison of the time course between the ERP waveform (at electrode FCz) of the Load 5 – First set presentation condition of the mixed experiment (Experiment 1) and the activations of IC2.

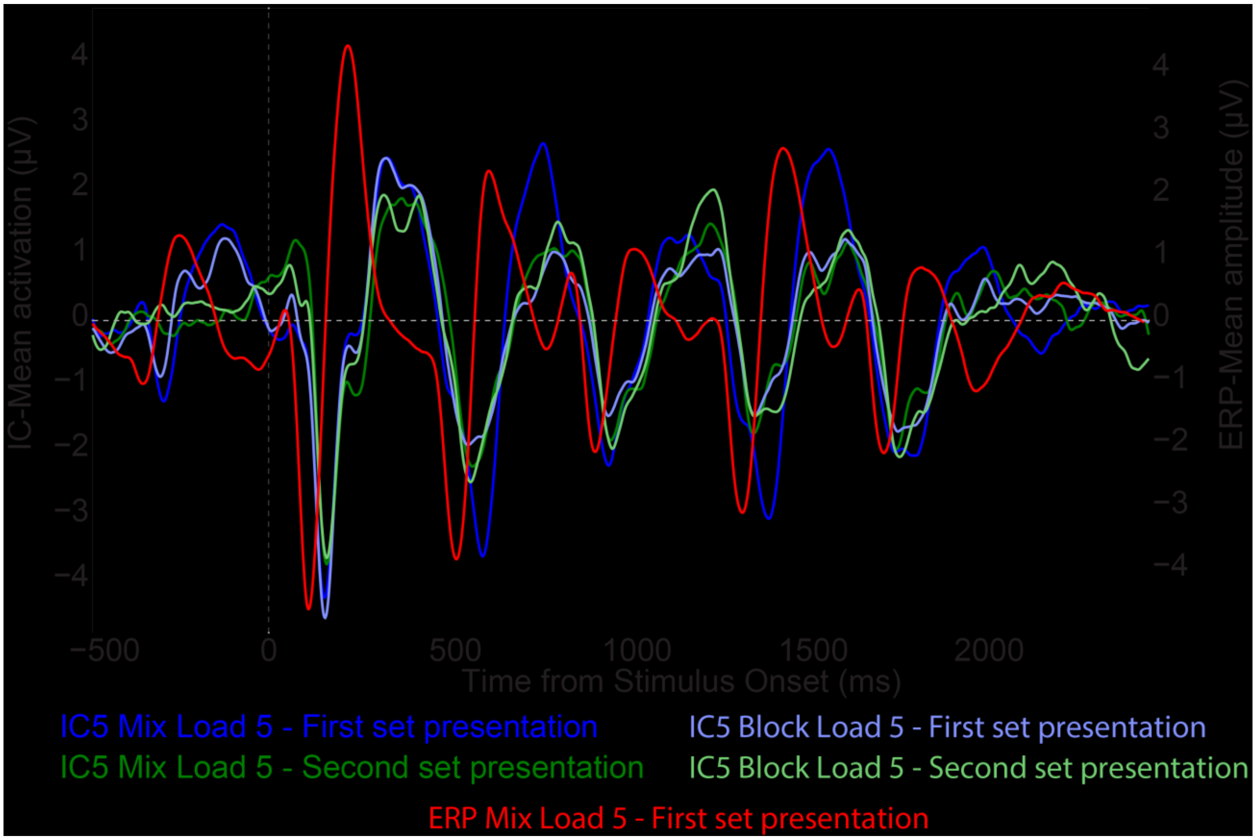

Figure 2.S4. Comparison of the time course between the ERP waveform (at electrode FCz) of the Load 5 – First set presentation condition of the mixed experiment (Experiment 1) and the activations of IC5.

Chapitre 3 – Oscillations linked to auditory informativeness and context disambiguation

---

Amour Simal[1,2,3], Robert Zatorre[2,4], & Pierre Jolicœur[1,2,3]

1. Département de psychologie, Université de Montréal, Montréal, QC, Canada

2. Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montréal, QC, Canada

3. Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, QC, Canada

4. McGill University–Montreal Neurological Institute, Neuropsychology and Cognitive Neuroscience, Montreal, Canada

3.1 Abstract

Accurate predictions and the processing of prediction error signals can be important for efficient interaction with the auditory environment. In a reanalysis of data from Simal et al. (2021), we sought to identify new electrophysiological indicators associated with disambiguation of the hearing context and prediction of forthcoming stimulation. Participants heard two isochronous sequences of pure tones separated by a silent retention interval. A sequence could contain one, three, or five tones. Fifteen participants heard the three load conditions randomly intermixed. In this case, when sequence length was unknown, the second and fourth tone during encoding contained information allowing the prediction of another tone. Other participants heard the sequences blocked by sequence length, and the second and fourth tone of the sequences provided no new information (and hence were not informative). Informative tones elicited increased N1 and P2 event-related potential (ERP) components. Using wavelet analysis, we found a significant increase in theta (4 – 7 Hz) amplitude following a tone that was informative and allowed prediction, in comparison with a tone that carried no predictive information. Previous work suggests increased theta amplitude is linked with task switching and an increase in cognitive control. We suggest informative tones recruit higher-level control processes involved in prediction of upcoming auditory events. Alpha (8 – 10 Hz), however, increased during encoding, but was not modulated by tone informativeness.

3.2 Introduction

Predictions and prediction error signal are of crucial importance for us to interact efficiently with our environment. This is particularly true in the auditory domain in which the transient nature of the stimuli makes the accurate deployment of attention imperative. A recent line of work has shown electrophysiological indices related to the disambiguation of the hearing context and the generation of predictions (Simal et al., 2021; Chapter 2 of this thesis), namely an increase in the amplitude of N1 and P2 ERP components.

In this study, we analysed this dataset in the time-frequency domain in order to understand better the oscillatory mechanisms involved. Because, to our knowledge, no results similar to those described by our laboratory (Simal et al., 2021) have been examined by another group, nor by observing responses in the time frequency domain, we could not make strong a priori predictions for which frequency bands might be affected, how, or when, following the presentation of informative tones. We used a wavelet transformation to explore the dynamics of oscillatory brain activity in our paradigm (Pantazis et al., 2009; Tallon-Baudry et al., 1997; Tallon-Baudry & Bertrand, 1999).

Simal et al. (2021) performed two experiments in which sequences of tones were presented in the context of a simple memory task. Each trial had three phases: encoding, retention, test. In both experiments, sequences containing one, three, or five tones to be remembered were presented. After a silent retention interval of 2 sec, a test sequence containing the same number of tones as the first was presented. The task was to decide if the two sequences were the same or different (when different, two adjacent tones were presented in reverse order; for trials with a single tone, the tone was replaced). In the first experiment (Mixed experiment), sequence length was random from trial to trial so participants did not know how many tones they would hear during encoding. In this context, a second or fourth tone, if presented, was informative because one more tone would always follow (sequences of length two or four never occurred). In the test sequence, however, sequence length was the same as for encoding, and thus known is advance. In this case, tones two or four provided no new information about sequence length. In a second experiment (Blocked experiment), the three load conditions were presented in different blocks, so for about 200 trials, participants were hearing the same number of tones in each sequence, and no tones were disambiguating, even in the encoding phase.

Our hypothesis was that, following the results presented by Simal et al., oscillatory modulations related to the generation of predictions and disambiguation of hearing context, should be present in the mixed experiment during the encoding phase, for the second and fourth tone, but not for any other tone of any other sequence.

3.3 Methods

The dataset is the same as described in (Simal et al., 2021), also refer to Chapter 1 of this thesis.

3.3.1 EEG analyses

EEG signals were sampled continuously at 256 Hz with a 72-channel BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes within an elastic cap with placements that followed the International 10–10 system (Sharbrough et al., 1991). Additional electrodes were placed at the left and right outer canthi to track horizontal eye movements, below the left eye to track vertical eye movements (in combination with the frontal electrode Fp1 that is placed above the left eye), and on the left and right mastoids. EEG data were re-referenced offline to the average of the two mastoids. The difference between the electrodes on the left and right outer canthi was defined as the horizontal electrooculogram (HEOG), and the difference between the electrode below the left eye and Fp1 as the vertical eletrooculogram (VEOG).

EEG data was analyzed using EEGLAB (Delorme & Makeig, 2004), ERPLAB (Lopez-Calderon & Luck, 2014), Brainstorm (Tadel et al., 2011), and custom Matlab code. An Independent Component Analysis (ICA, see Drisdelle, Aubin, & Jolicœur, (2017)) was performed on the EEG from each subject and components corresponding to ocular artefacts were removed from the data, allowing the removal of saccades and blinks. Ocular artefacts were detected by the use of an algorithm that correlated ICA components to a template of a blink or a saccade and by correlations of the component activation time courses to the VEOG and HEOG observed difference waves.

Afterwards, a high-pass filter of 0.05 Hz and a low-pass filter of 128 Hz was applied, using ERPLAB scripts and using a Butterworth filter (zero-phase), 6th order, appropriate to eliminate slow waves, likely to be artefacts, as well as unusable data in the time frequency domain according to Nyquist theorem. We also used a notch filter to remove activity around 60Hz, the frequency in the electrical system in Canada. The EOG signals (HEOG and VEOG difference waves) were filtered with a 0.1

Hz high-pass filter and a 10 Hz low-pass filter. EEG was epoched from 500 ms before the first tone of the set to 2,500 ms after the start of set presentation, for each set. This meant epochs ended 700 ms after the last tone of each set for the Load 5 condition. For the second set presentation, we only kept trials for which both sets were the same, such as not to include effects of tone deviation, or cognitive responses related to response selection that could be confounded with possible effects of expectation. We will only focus on the Load 5 condition because it allows us to perform the most relevant comparisons of effects between tones (and interactions with of tones and experiments and phases of the memory task).

Before doing time frequency analysis, we removed the trials containing artefacts. Ocular artefacts[6] were detected if the HEOG signal varied by more than 35 μV over a 300-ms period, and/or if the VEOG signal fluctuated by more than 50 μV over a 150-ms period. Scalp-EEG electrodes with non-eye movement-related artifacts in the EEG signals (signal exceeding ±100 μV) were interpolated (at the exception of FCz, the electrode of interest), only if seven channels or fewer needed interpolation, otherwise the trial was excluded from further analysis.

3.3.2 Wavelet analysis

To explore the effects of tones informativeness without being restricted to a frequency band, the data was convolved with a complex Morlet wavelet from 1 to 50 Hz in 1 Hz increment. Wavelets were formed from a mother wavelet of (central) frequency 1 Hz and a time resolution of 1s (full width half maximum) at that central frequency. These parameters allowed a good spectral and temporal resolution at frequencies of interest in theta and alpha bands (see Tadel et al., 2011, for further details of the properties of this wavelet). The time-frequency data for each subject was normalized by baseline correction, and transformed to reflect event-related changes in oscillatory power. A value of 0% indicates the signal was at baseline, positive values indicate an increase in power, and negative values indicate a decrease in power the signal.

---

[6] Artefact remaining on the electrooculogram that are either ocular artefact that could not be corrected by the ICA or non-ocular artefacts.

3.4 Results

3.4.1 Wavelet results

The wavelet analysis results at electrode of interest FCz can be seen in Figure 3.1. Visual inspection shows alpha throughout presentation of the first sequence (encoding), with apparent augmentation of power following informative tones (first sequence, mixed experiment). Because peaks seem to be in the lower alpha range, our subsequent analysis will focus on the 8 – 10 Hz frequency range. We can also observe theta power increase after informative tones (around 5 Hz), starting around 100 ms after onset of informative tones (2nd and 4th tone, indicated by orange bars in Figure 3.1), and lasting approximately 300 ms.

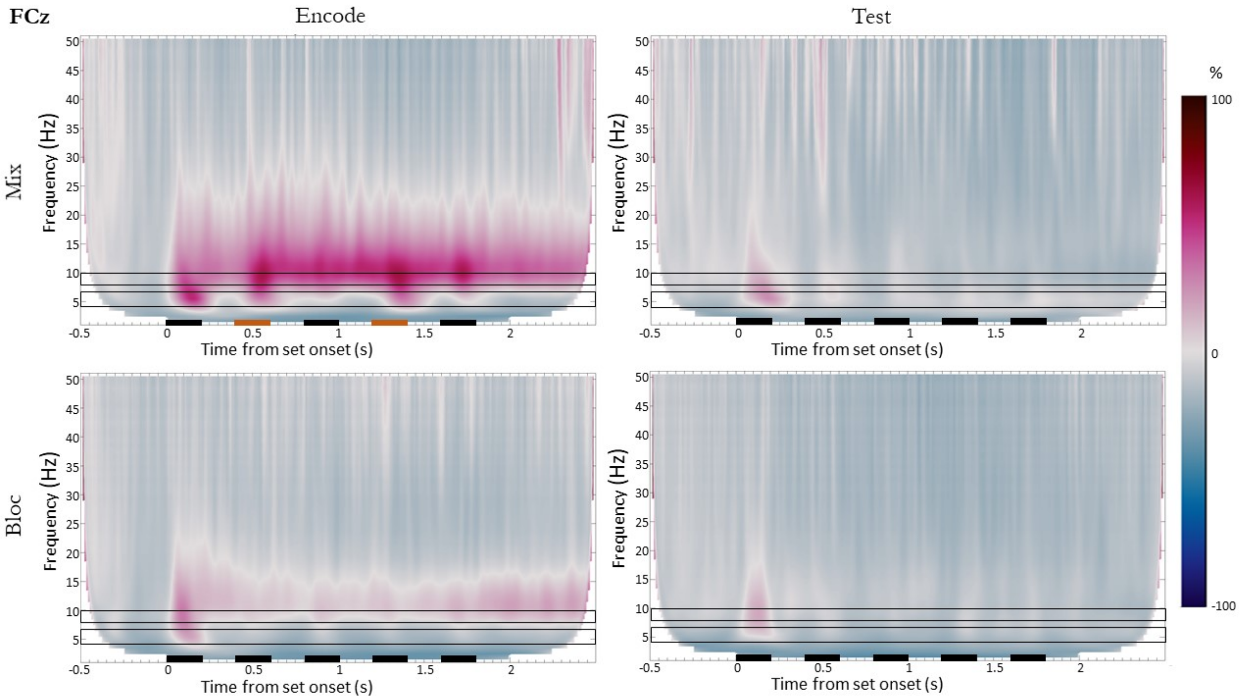

Figure 3.1. Event-related time-frequency maps at electrode FCz for each set presentation of each experiment. Oscillatory power (abs(TF)^2) is shown as a percentage increase or decrease relative to the mean power during a baseline of 500 ms before the onset of the first tone. Pink indicates an increase in power relative to baseline, and blue indicates a decrease. Tone presentations are represented by black bars (non-informative tones), or orange bars (informative tones).

Frequency windows for Hilbert transform analysis are highlighted with black boxes. Alpha was analyzed between 8 and 10 Hz (low alpha), and theta between 4 and 7 Hz.

3.4.2 Hilbert transform

To quantify activity in alpha and theta band we used the Hilbert transform, which gives the same information as the wavelet analysis in a different format (Bruns, 2004). We were interested to verify if any oscillatory activity was associated with tone presentation, and if it was stronger for informative tones, specially at electrode FCz. To extract the activity of interest for visualization, we subtracted the activity of the second set presentation from the activity of the first set presentation, for each experiment. We then subtracted the difference waves of the blocked experiment from the difference waves of the mixed experiment. Theoretically, this manipulation of the data eliminates activity related to low-level sensory effects from tone presentation, as well as memory-related activity. Figure 3.2 shows the original and differential time-frequency activity in the low alpha (8 – 10 Hz) and theta (4 – 7 Hz) range, as well as scalp plot associated with activity of interest. Visual inspection shows alpha increase for the first set presentation, throughout sequence presentation consistent with what is seen in Figure 3.1. Increases in alpha amplitude were distributed widely over scalp electrodes with higher effects at fronto-central (FCz) and parieto-occipital (POz) electrodes. Examination of oscillatory activity at electrode POz showed the effects consisted of an increasing alpha amplitude throughout sequence presentation during encoding, somewhat larger for the mixed experiment (see Figure 3.S1). We can also observe theta augmentation after informative tones, starting around 100 ms after onset of informative tones (2nd and 4th tone, indicated by orange bars in Figure 3.1 and Figure 3.2), and lasting approximately 300 ms.

Because the first tone presented (no matter in which phase) proved to have a specific electrophysiological signature that was equivalent in all sequences and hence provided no differential information, we did not include it in our analyses. The measured power change for each participant for the tone positions 2, 3, 4, and 5, were submitted to ANOVAs that included a specific contrast designed to test the hypothesis that the second and fourth tones would be different from the third and fifth, as in the analyses used by Simal et al. (2021). This pattern was encoded in the contrast weights (+1 -1 +1 -1) over tone positions.

We measured alpha amplitude from 100 ms to 200 ms following tone onset, spanning the end of the N1 to the beginning of the P2 component. There was no Experiment x Set x Contrast interaction, *F* < 1. The pattern in amplitudes across tones also did not differ between experiments for the encoding set (Experiment x Contrast), *F* < 1, nor the retrieval set (Experiment x Contrast), *F* < 1. The only significant effects were the differences between set, regardless of Contrast, for the mixed experiment, *F*(1, 14) = 8.64, *p* = .01, and the blocked experiment , *F*(1, 14) = 5.71, *p* = .03.

Theta amplitude was computed between 4 and 7 Hz and is shown in Figure 3.2, Panel B. Measurements were taken at electrode FCz from 100 to 250 ms after each tone onset, corresponding approximately to FWHM time windows. The was a significant Experiment x Set x Contrast, *F*(1, 28) = 21.99, *p* < .001. This was caused by the pattern of amplitude change between tones being different between sets for the mixed experiment (Set x Contrast), *F*(1,14) = 20.89, *p* < .001, but not for the blocked experiment, *F*(1,14) = 2.15, *p* = .16. In the mixed experiment, theta amplitude increased more in response to informative tones, than in response to non-informative tones, in the first set, *F*(1, 14) = 31.45, *p* < .001. This was not the case when tones were not disambiguating, in the second set of the mixed experiment, *F*(1, 14) = 3.05, *p* = .10, the first set of the blocked experiment, *F* < 1, nor the second set of the blocked experiment, F(1, 14) = 2.94, *p* = .11.

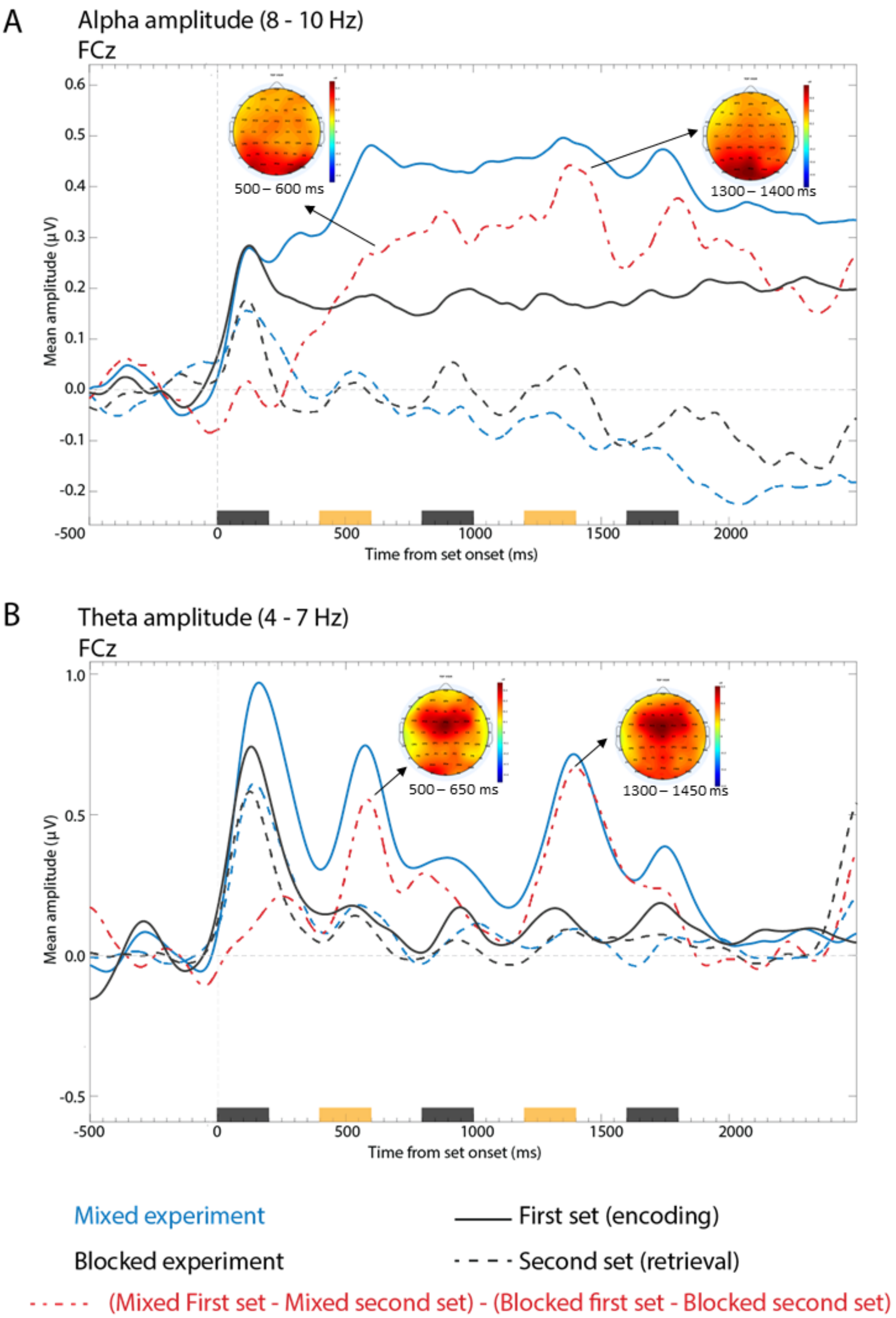

Figure 3.2. Oscillatory amplitude at electrode FCz. A) Low alpha (8 – 10 Hz) relative to baseline. B) Theta (4 – 7 Hz) relative to baseline. Boxes on the abscissa axis represent tone presentation, with yellow boxes indicating informative tones (only applicable to mixed Experiment, first set (encoding)).

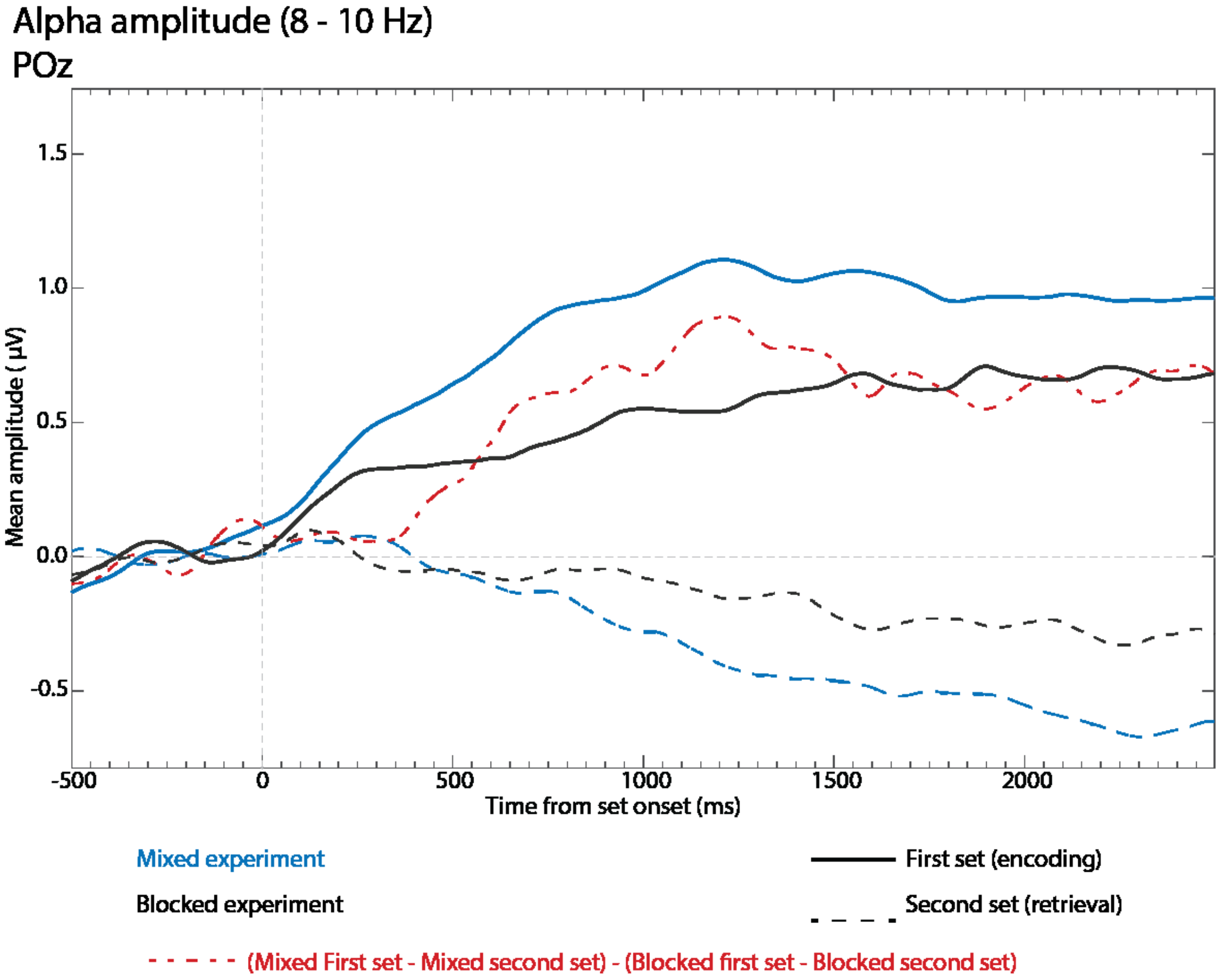

Figure 3.S1. Low alpha (8 – 10 Hz) amplitude at electrode POz for each set presentation of each experiment, as well as difference wave isolating activity specifically related to Mixed experiment First set (encoding).

3.5 Discussion

Our results show an oscillatory signature associated with tones disambiguating sequence length. Theta power was significantly related to tone informativeness and showed a phasic, rather than tonic, increase at frontocentral electrodes when a prediction could be made. Maximum increase in theta amplitude was observable over frontocentral electrodes. Although low alpha in the wavelet analysis showed hint of modulation by informative tones, the statistical analysis demonstrated this was not the case (see analyses based on the Hilbert transform).

In the frequency domain, low frequency modulations, such as in the theta band, are related with changes in neuronal excitability and in attention/perception (Fries, 2015). Because of the scalp distribution, coherent with sources in the auditory cortices, it is possible the increase in theta reflected a change in neuronal excitability to process better subsequent stimuli. This could be the case in our experiment because the prediction generated a need for change in working memory buffer. An increase in theta related to task switching has also been reported (Sauseng et al., 2006). In some experiments, frontal midline theta has been shown to originate from mid-cingulate cortex and pre-supplemental motor areas and is thought to be involved in recruiting other brain areas to implement cognitive control in contexts involving uncertainty about action and outcomes (Cavanagh & Frank, 2014). In the mixed experiment, the presentation of more than one tone is uncertain, and it is possible the presentation of a second or a fourth tone creates a need for readjusting attention and goals (Danielmeier & Ullsperger, 2011), somewhat in line with the idea of task updating/switching. Although an increase in theta has been observed in MMN paradigms, showing sensitivity to violation or expectation and prediction error (Fuentemilla et al., 2008), it is unlikely what we describe involves the same processes and brain areas, because the resulting measured ERP has the opposite polarity, and that the increase in oscillatory amplitude they describe range from theta to low and high alpha. Because of the frontocentral/frontal midline scalp distribution, it is not possible, with the current dataset, to determine if theta augmentation originates in the auditory cortices, or cingulate cortex and preSMA, or both. More research will be needed to resolve this question.

The global increase in alpha amplitude is in line with working memory studies showing alpha increase during encoding in auditory working memory tasks (Krause et al., 1996; Pesonen et al., 2006). Because alpha modulation had a wide role in attention and cognitive processe (Başar et al., 2001; Palva & Palva, 2007), and is thought to be linked to top-down inhibitory control (Klimesch et al., 2007), it is not surprising we did not find phasic activity in response to a tone allowing prediction.

In conclusion, our results suggest that a tone disambiguating the context and allowing prediction to be made recruits higher level brain areas and this activity is reflected in higher theta oscillations.

Chapitre 4 – Detecting auditory temporal regularities: electrophysiological index of tracking and identification of disambiguating information

---

Amour Simal[1,2,3], Robert Zatorre[2,4], & Pierre Jolicœur[1,2,3]

1. Département de psychologie, Université de Montréal, Montréal, QC, Canada

2. Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montréal, QC, Canada

3. Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, QC, Canada

4. McGill University–Montreal Neurological Institute, Neuropsychology and Cognitive Neuroscience, Montreal, Canada

4.1 Abstract

Learning, and detection of regularities allows us to make predictions about our environment and process stimuli more efficiently. Using EEG, we found an electrophysiological signature linked to how the brain uses and interprets auditory information in the time domain. We used sequences of five tones with different pitches, with one of three distinct rhythmic pattern (RP): short-long-short-long, long-short-long-short, or isochronous ISI patterns. They were designed so that the second tone always carried temporal-sequence information, by being presented after a short, medium, or long ISI, allowing recognition of the pattern. On each trial participants heard two tone sequences with the same RP and had to indicate if the tone pitches were identical. In one experiment, the three types of RP were randomly intermixed, whereas they were blocked in a separate control experiment with a new group of participants. A frontal and frontocentral positivity increased for the encoding sequence of the first experiment (when RP was not previously known), compared to that same sequence presentation in the control experiment (pattern known), starting around the earliest time the second tone could be presented, and peaking shortly after actual tone onset. Although these temporal patterns were task irrelevant, and most participants were unaware of them when asked, our results suggest that the brain disambiguates its variable environment based on the earliest available information, and that it does so rapidly, pre-attentively, and automatically.

4.2 Introduction

The ability to predict changes in our environment is a crucial function that allows us to direct our attention to relevant items while ignoring all superfluous information. Predictive processes have been the subject of numerous studies, both on a behavioral, and on a neuroimaging level. Most of the research has been focused on brain response to stimuli violating or confirming predictions such as the mismatch negativity (MMN; May & Tiitinen, 2010; Näätänen et al., 2007, 2011) or the repetition positivity (RP; Haenschel et al., 2005).

Despite decades of research, debates about the cognitive processes underlying the MMN remain, such as whether it is generated by existing predictions, or post-hoc (Denham & Winkler, 2020). In the first case, the system extracts regularities from the environment and uses them to make predictions. Each input is then compared with the prediction, and in this case the MMN is a mismatch between the prediction and the stimulus (Winkler, 2007). The post-hoc processes, would be based on memory traces or adaptation (P. J. C. May & Tiitinen, 2010; Näätänen, 1990). One of the strongest pieces of evidence to date in support of the predictive model are the results from Bendixen et al. (2009) study that showed the EEG signal to a predictable, but omitted sound, is undistinguishable from the signal of the actual sound presentation in the first 50 ms. It was argued that, to respond to an absent sound, the brain must have generated a prediction. One caveat to this result however is that these omission responses are only present with an SOA smaller than 170 ms (Yabe et al., 1998). Some studies have shown prior knowledge to have no effect on the MMN (Sussman et al., 2003), that is, the MMN is elicited even though the 'deviant' stimulus is predictable (by visual-auditory link), supporting the memory trace view. However in other studies (e.g., Sussman et al., 2002), explicit knowledge about sequence structure eliminates the MMN response, supporting the predictive view. A recent study by Tsogli et al. (2019) highlighted that both processes are at play. They refer to the memory trace MMN as the physical MMN, caused by changes in low-level physical properties of the sound, and to the prediction-related MMN as the statistical MMN, modulated by transitional probabilities. Their results showed the statistical MMN interacts with the physical MMN in that the prediction error of low-level physical attributes takes precedence over learned statistical properties and suppresses the statistical MMN.

We aim to add evidence for predictive coding to the debate. However, instead of studying responses to a stimulus matching or mismatching predictions, we want to examine responses to an event that allows predictions to be generated by a stimulus that disambiguates the hearing context. A previous study from our laboratory showed a specific neural signature for tones that disambiguate sequence length (Simal et al., 2021). In a context in which the lengths of tone sequences were one, three, or five tones long, at random, (but never two or four) the presentation of a second or fourth tone indicated at least one more tone would be heard. A second or fourth tone (informative in this context) generated a different pattern of brain activity from the other tones (expected tones, 1, 3, 5), or than the same tones in a context in which sequence length was known beforehand. The difference was reflected in a larger P2, as well as, to some extent, a larger N1. Simal et al. argued in favor of a predictive process, because the modulations of the ERPs could not have been caused by low-level effects such as adaptation or memory traces. The response to tones that were informative of other tones to come occurred only in the context of uncertainty concerning sequence length.

4.2.1 Temporal predictions

Instead of examining an event disambiguating sequence length, demonstrated by Simal et al. (2021), here we were interested in events disambiguating the temporal structure, or rhythmic pattern, of a sequence. In the present paper the terms pattern, rhythmic pattern, or RP will refer to a combination of different ISIs in a sound sequence. In a more general context, temporal structure could include the duration of individual auditory events, but we will not examine those effects in this paper.

The recognition of the rhythm, or rhythmic pattern, of a song is largely independent from the tempo at which it is presented. Two melodies are perceived as the same, but transposed, if one is obtained by a multiplicative transformation of the other (Hulse et al., 1992). For temporal structure, “perceptually invariant” by Hulse et al.’s definition does not mean the melodies are indistinguishable, but that the structure is recognizably the same. This means the relative temporal relationship between auditory elements is more important than the absolute relationship. A multiplicative transformation maintains this structure, and the modified stimulus tends to be perceived as the same as the original stimulus. A recent study by Notter et al. (2019) showed the brain encodes relative rhythmic patterns as temporal Gestalts, regardless of their absolute timing. They provided evidence for the involvement of frontal and temporoparietal regions, also involved in melodic transposition, pitch reversal (Foster et al., 2013) and visual Gestalts (Huberle & Karnath, 2012). These results, and others (e.g., Belin et al., 2002; Chen et al., 2008; Geiser et al., 2012; Grahn & Rowe, 2009), suggest encoding and integration of rhythmic patterns as Gestalt representations involves higher-level associative areas of the brain. We hypothesize these representations could also be a basis for predictions and disambiguation of auditory context.

Our goal was to discover an electrophysiological signature of temporal pattern recognition. We created a paradigm with simplified rhythmic patterns that allowed us to control at which point the recognition of the pattern could occur. It was important to ensure participants paid attention to the auditory stimuli, though not necessarily to temporal patterning. To do this we embedded the different rhythmic patterns in the context of an auditory memory task in which participants had to encode the pitch of tones in a first sequence of tones (encoding), followed by a short retention interval, and finally a second sequence of tones that had the same pitches or differed in the pitch of one of the tones. As in the experiments of Simal et al. (2021), the task was to decide whether the two sequences were the same or different (in pitch). We created three rhythmic patterns in sequences containing 5 tones with the constraints that sequences should have equal total duration, that the moment of presentation of the second tone would be informative, and that the patterns following the second tone should all differ. The three different patterns are shown in Figure 4.1. One of the three condition was isochronous sequences, called RP1. In this condition, all ISI have the same length (300 ms, "medium ISI"). The second tone could also be presented at a short (RP2), or long interval (RP3) following the first tone. If the interval was shorter, the following intervals were long, short, long, creating a possibility of grouping tones 1 and 2, and tones 3 and 4, which we called RP2. The third pattern (RP3) was special because we expected the second tone itself might no longer be informative, given information of the time elapsed since the first tone. If the brain was tracking the second tone presentation, by not hearing it in a short or medium interval, then the tone would be, by default, presented at the longer interval. In RP3, the ISIs were long, short, long, short. In the design of this experiment, RP3 served as a "placeholder" condition that could be inferred by not having heard the second tone shortly after the middle ISI.

We used two versions of the experiment, similar to the experimental protocol used by Simal et al. (2021). In Experiment 1, also referred to as the "mixed experiment," we presented a randomly chosen rhythmic pattern on each trial. Thus, the temporal structure of the initial encoding sequence was not known at the beginning of the trial. The presentation of the second tone was then sufficient to disambiguate which rhythmic pattern was used in that trial. The second tone should have no special status during the test sequence (retrieval), however, because the rhythmic pattern was always the same as in the encoding sequence. In order to control for differences in memory-related processes that differ between sequences, we also performed a separate control experiment with a separate set of listeners. Experiment 2, referred to as the "blocked experiment," was identical to Experiment 1 except that the rhythmic patterns were presented in different blocks of trials. Thus, the rhythmic pattern was known in advance even in the encoding sequence. In this case the second tone was not diagnostic or disambiguating and we expected the first sequence (encoding) of the blocked experiment, acting as a control, to differ from the first sequence (encoding) of the mixed experiment, while the second sequence (test) should yield similar results between experiments.

4.3 Method

4.3.1 Participants

Participants were recruited by announcements at Université de Montréal and via social media. They gave informed consent to procedures vetted by the local institutional bio-imaging ethics committee. Protocols regarding mitigation of COVID-19 were followed. All participants received financial compensation of 20 CAD. All participants reported having no history of neurological disorder and having normal hearing. We verified their hearing with an audiogram and did not test participants with a threshold higher than 25 dB SL in frequencies between 0.5 kHz and 4 kHz.

Twenty-five participants were recruited in Experiment 1. The data from one participant could not be used because the EEG data file was corrupted and did not contain all the event markers. The remaining 24 participants were between 18 and 37 years old (M = 22.1, S.D. = 4.3), 18 women, 3 left-handed.

Twenty-six participants were recruited in Experiment 2. The data from one participant was rejected because response accuracy was close to chance (54.4%); and data for another participant was accidentally not recorded. Remaining 24 participants were between 18 and 35 years old (M= 23.7, S.D.= 4.1), 17 women, 3 left-handed. Four participants had previously participated in Experiment 1.

4.3.2 Stimuli

Tones used in the experiment were 14 non-musical pure tones not corresponding to notes in the well-tempered scale. Their exact frequencies were 380, 419, 463, 511, 564, 623, 688, 759, 838, 925, 1022, 1128, 1245, and 1375 Hz. They were obtained by multiplying the starting frequency (380Hz), as well as each resulting frequency by $2^{1/7}$. This means differences between the tone were equal on a log scale, with each octave divided into seven equal steps. Tones had a duration of 200 ms (including 10 ms rise and fall times). Tones were generated using Audacity®[7] at 16-bit depth with a sampling rate of 44.1 kHz. Stimuli were presented binaurally at 65 dB SPL (adjusted for the 1022 Hz tone) via Etymotics ER-2 earphones and foam inserts.

---

[7] Audacity® software is copyright © 1999-2021 Audacity Team.
Web site: https://audacityteam.org/. It is free software
distributed under the terms of the GNU General Public License.
The name Audacity® is a registered trademark

Tone sequence structures were made with the aim of having the second tone be disambiguating of temporal pattern with the constraint that all sequences contain the same number of tones and had the same duration. The three types of temporal pattern are illustrated in Figure 4.1.

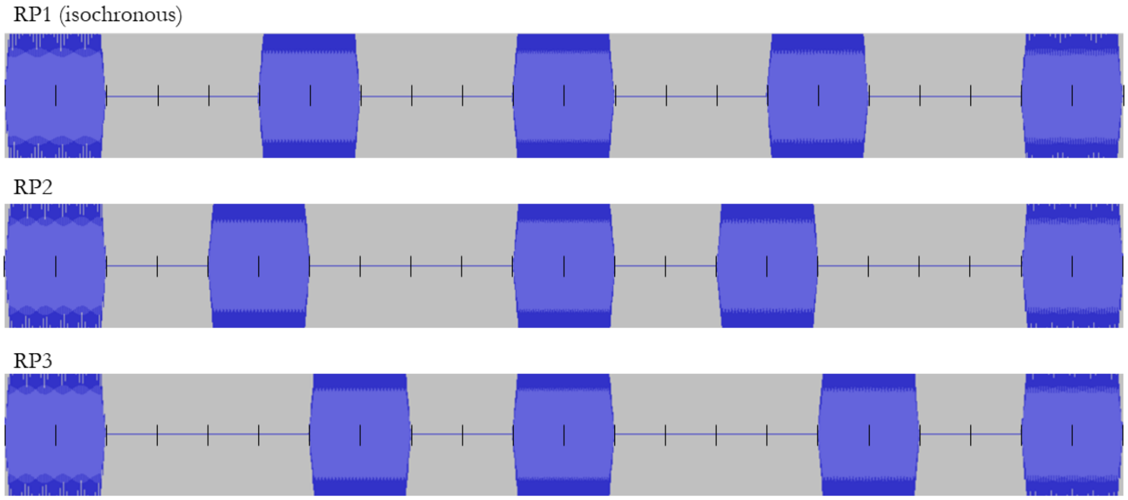

Figure 4.1 Three types of rhythmic patterns in the context of the experiment. Tones were 200 ms pure tones with 10 ms rise and fall. Black tick marks are spaced 100 ms apart. The second tone presented is disambiguating of the temporal presentation of the rest of the sequence when sequence pattern is not known prior to presentation.

4.3.3 Experiment 1 (Mixed presentation)

Participants were in a dimly-lit Faraday room in front of a computer screen. Eighteen participants were in a larger room, and sat in a chair at a distance of 97 cm from the computer screen with a keyboard on their lap. The other 6 participants were in another smaller room and were seated in front of a table with the computer screen at 57 cm. The size of visual stimuli (fixation cross, response disk, and feedback) on the screen was adjusted to equate visual angles across the two viewing distances.

Participants performed a simple memory task with three phases: encoding, retention, test. Each trial started automatically with the appearance of a grey fixation cross on a black background.

After a delay of 500 ms plus or minus a random jitter of up to 100 ms, we presented the memory encoding sequence, that was a random sequence of 5 unique tones chosen randomly without replacement from the set of 14 possible tones. The rhythmic pattern, determining the time intervals between tones, was chosen randomly from the three possible rhythmic patterns shown in Figure 4.1. After a silent 2000 ms retention interval, they were presented with a second sequence, the test sequence, that always had the same rhythmic pattern as in the encoding sequence. The tones in the test sequence either were the same (the same pitch, in the same order) as in the encoding sequence, or the pitch of one of the tones was changed to a pitch that was not presented in the encoding sequence. Participant waited for 500 ms after the end of the tone sequence, when fixation cross changed to a white disk, and then indicated whether the sequences were the same or different using the keyboard. Response keys were "F" and "J" on a QWERTY keyboard and were counterbalanced between participants. After a mean delay of 500 ms, with a random jitter up to ± 100 ms (between 400 and 600 ms), the white disk on the screen turned green for a correct answer, red for an incorrect answer, or deep red if participants did not answer within 3000 ms (time-out). This feedback was presented for 300 ms. The screen became black for 2000 ms before the start of the subsequent trial. The trials were presented in 8 blocks of 42 trials. Each block contained seven occurrences of each condition (6 combinations of 3 temporal patterns and same vs. different pitches). The first two trial blocks were not included in the analyses to allow implicit learning of the rhythmic patterns. We analysed a total of 84 trials per rhythmic pattern condition. A practice block containing 4 trials was presented at the beginning of the experiment to ensure participants understood the task and all equipment was functional. Accuracy feedback was provided on the screen at the end of each block to the participant and the experimenter to keep track of performance and to incentivise participants to improve their score and keep them motivated.

4.3.4 Experiment 2 (Blocked presentation)

Participants were tested in the smaller room described in Experiment 1. The experiment was identical to Experiment 1, except for the following differences. Firstly, the rhythmic pattern conditions were blocked in Experiment 2 but mixed from trial to trial in Experiment 1. Secondly, the trials were presented in 9 experimental blocks containing 36 trials. Each rhythmic pattern condition was presented in 3 consecutive blocks, for a total of 108 trials per condition, with a practice block containing 4 trials preceding each new rhythmic pattern condition.

4.3.5 EEG

In both experiments, EEG was recorded with a 72 channel BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes, at placements following the international 10–10 system (Sharbrough et al., 1991), at a sampling rate of 512 Hz. Additional electrodes were placed at the left and right mastoids (later used as references), at the left and right outer canthi (later used for horizontal electrooculogram, HEOG), and under the left eye (later used for vertical electrooculogram, VEOG, with electrode Fp1).

Data were analyzed using EEGLAB (Delorme & Makeig, 2004), ERPLAB (Lopez-Calderon & Luck, 2014), as well as custom Matlab code. Ocular artefacts were removed from the data with the use of an Independent Component Analysis using the method described in Drisdelle et al. (2017). Data were then filtered with a high pass of 0.5 Hz and a low pass of 30 Hz Butterworth filter. EEG was epoched from 200 ms before the start of tone sequence presentation to 2800 ms after the start of sequence presentation (400 ms after the offset of the last tone). Baseline correction was applied by computing, and removing from each epoch, the average amplitude in the 200 ms preceding start of tone sequence. To avoid any activity related to response selection or elicited by a mismatching tone, for the test sequence presentation, we only kept trials in which tones were identical to those in the encoding sequence. Artefact rejection was then applied, first by detecting ocular artefacts not corrected by the ICA. Trials during which HEOG varied by more than 35 μV over a 300 ms period, and/or during which the VEOG varied by more than 50 μV over a 150 ms period were removed. Remaining non-ocular artefacts were detected when signal exceeded a variation of 100 μV. If less than 7 channels contained an artefact, they were interpolated, otherwise the trial was rejected.

4.4 Results

4.4.1 Behavioral

For Experiment 1 (mixed rhythmic patterns), mean accuracy was 83.5% (*SD* = 6.9%). Shapiro-Wilk normality test indicated the distribution of scores across participants was not significantly different from a normal distribution ($W = 0.96, p = .35$). The mean music experience in this group was 5.5 years of music training, ranging from 0 to 20 years (M = 5.5 years, SD = 5.9). Music training was found to be non-normally distributed ($W = 0.85, p = .002$). Correlation between accuracy and years of music experience was significant, according to Kendall's tau correlation coefficient (τ = 0.32, *p* = .03, see Figure 4.S1).

In Experiment 2 (blocked rhythmic patterns), mean accuracy was 78.2% (*SD* = 7.7%). The distribution of accuracy scores across participants was visually similar to that observed in Experiment 1 (mixed rhythmic patterns). The distribution was found to be near-normal based on the Shapiro-Wilks normality test (*W* = 0.93, *p* = .088). The mean years of musical training in this group was 2.25 years, ranging from 0 to 10 years (*M* = 2.25 years, *SD* = 3.0). The distribution of music experience was non-normal (*W* = 0.75, *p* < .001). As shown in the scatter plot in Figure 4.S1, the correlation between music experience and accuracy was marginally significant, according to Kendall's tau correlation coefficient (τ = 0.30, *p* = .06). It is possible correlation did not reach significance because the range of music experience was more restricted than in Experiment 1.

Mean accuracy was significantly higher in Experiment 1 (mixed rhythmic patterns, *M* = 83.5%) than in Experiment 2 (blocked rhythmic patterns, *M* = 73.2), (*t*(45) = 2.52, *p* = .01), but participants in Experiment 1 had more musical experience (*t*(34) = 2.40, *p* = .02).

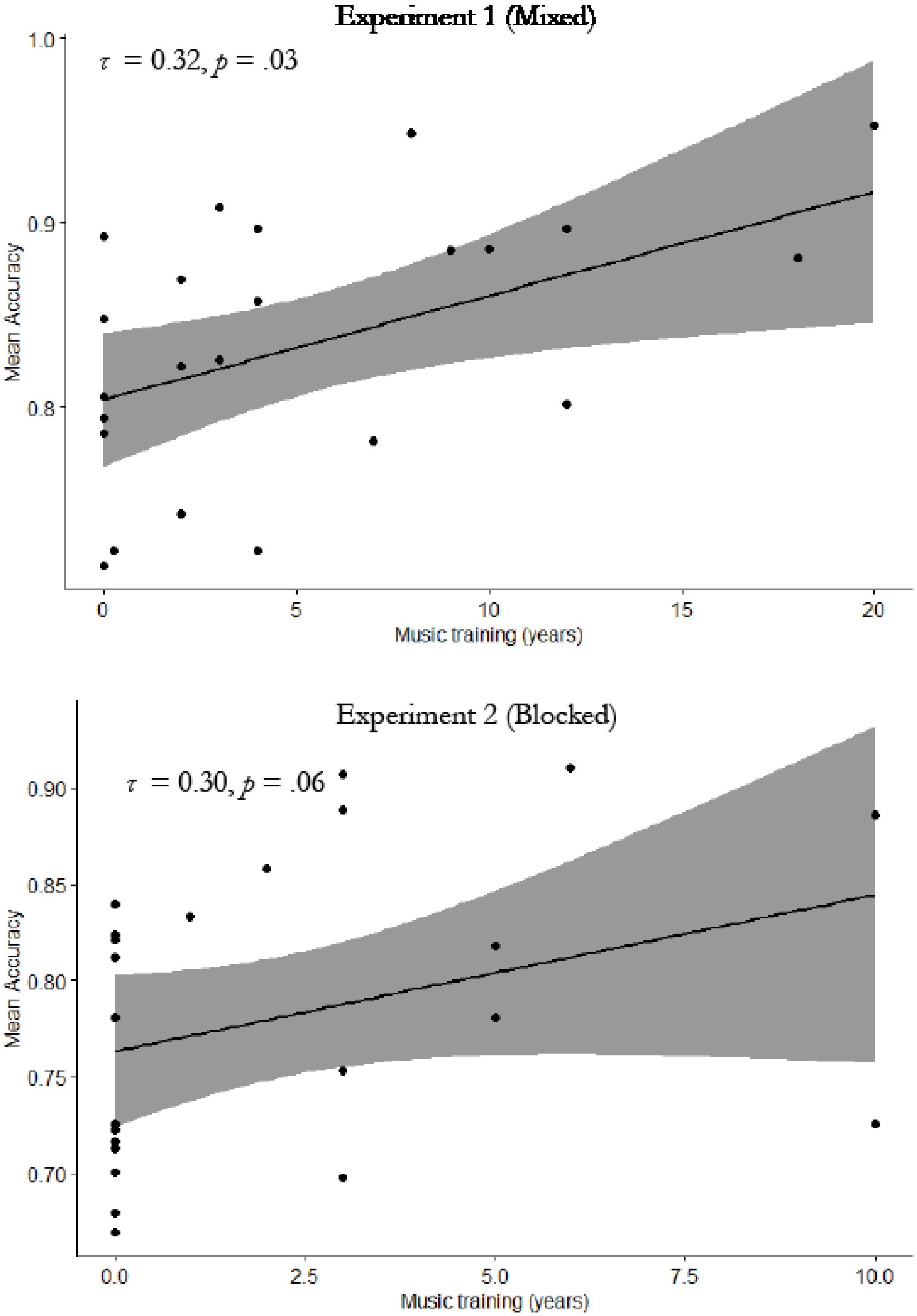

Figure 4.S1. Scatter plot of the mean accuracy and music experience of each participant for each experiment. Correlation was measured using Kendall’s rank test because the distribution of music experience was not normal. Linear regression is illustrated in the plot, as well as 95% confidence intervals.

4.4.2 EEG results

EEG analyses were performed including error trials. An analysis to compare results with or without error trials showed no difference. Since the results we describe hereafter are the same regardless, we made the decision of including error trials to keep a larger number of trials and to obtain a better SNR

Cortical activity measured from a pooling of the amplitude measured at fronto-central electrodes Cz, FCz, and Fz can be seen in Figure 4.2.

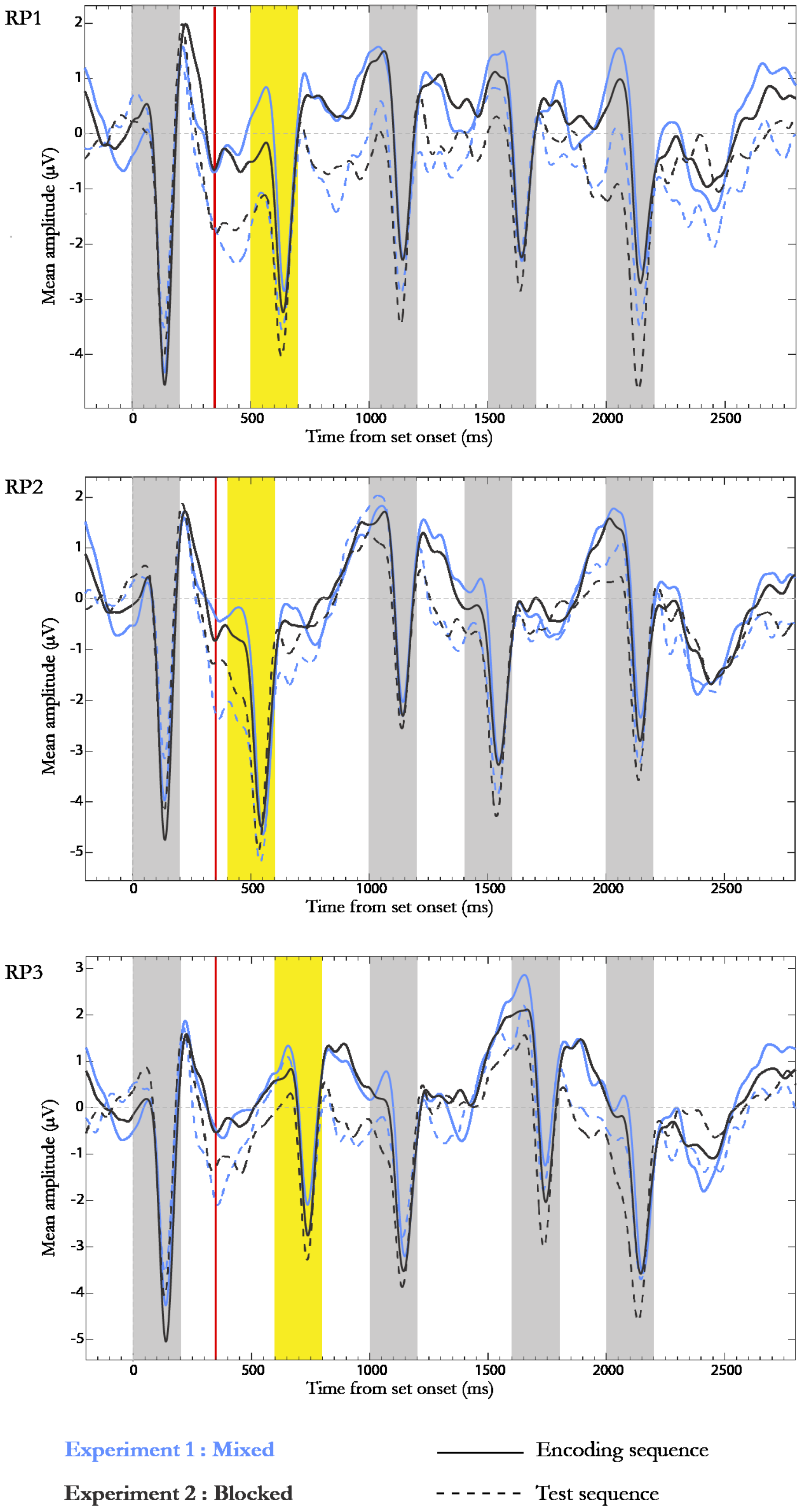
RP1
2
1
0
-1
-2
-3
-4
Mean amplitude (µV)
0
500
1000
1500
2000
2500
Time from set onset (ms)
RP2
2
1
0
-1
-2
-3
-4
-5
Mean amplitude (µV)
0
500
1000
1500
2000
2500
Time from set onset (ms)
RP3
3
2
1
0
-1
-2
-3
-4
-5
Mean amplitude (µV)
0
500
1000
1500
2000
2500
Time from set onset (ms)
Experiment 1 : Mixed
Experiment 2 : Blocked
Encoding sequence
Test sequence

Figure 4.2 ERP waveforms at an average of electrodes Cz, FCz, and Fz showing cortical activity from the onset of tone sequence. Each panel shows waveforms for encoding and test sequences, for each experiment (mixed vs. blocked rhythmic conditions). Tone presentations are indicated by a grey highlight. Tone of interest (second tone) is highlighted in yellow. Starting about 150 ms after the offset of the first tone, or 50 ms before the second tone would be presented in the shortest ISI condition (RP2), the waveform for the encoding sequence of the mixed experiment (solid blue line) starts to become more positive than the waveform for the encoding sequence of the blocked experiment (solid black line, control condition), timing indicated by a red line. Shortly following onset of the second tone, this positivity is maximal.

We were mostly interested in the activity in response to the second tone, which should be disambiguating of rhythmic pattern according to our hypothesis. In the waveforms, we can see a larger positivity when the tone was presented in an uncertain context, that is, in the encoding sequence in the mixed experiment, than when the rhythmic pattern was known beforehand, that is, in the encoding sequence in the blocked experiment or the test sequence in both experiments. In Figure 4.2, we highlighted, with a red line, when the activity for the encoding sequence diverged between the two experiments, based on visual inspection. It is around 150 ms following the offset of the first tone, or about 50 ms before the second tone would be presented in the shorted ISI condition (RP2). This would be coherent with the brain tracking the timing of tone presentation starting shortly before the informative tone could be presented (RP2), until actual presentation of the tone. This tracking is reflected by a positivity increasing with time and peaking right after the onset of the second tone. For the statistical analysis, we measured ERP activity peaking 50 ms post-stimulus presentation, using a window 60 ms wide (+20 to +80 ms relative to tone onset). Scalp topographies of the differential activity between encoding and test sequences during that time window can be seen in Figure 4.3.

The negativity over parietal /occipital electrode sites seems related to encoding-related processes, as it reflected a long-lasting negative-going wave for the encoding sequence regardless of experiment This will not be discussed further.

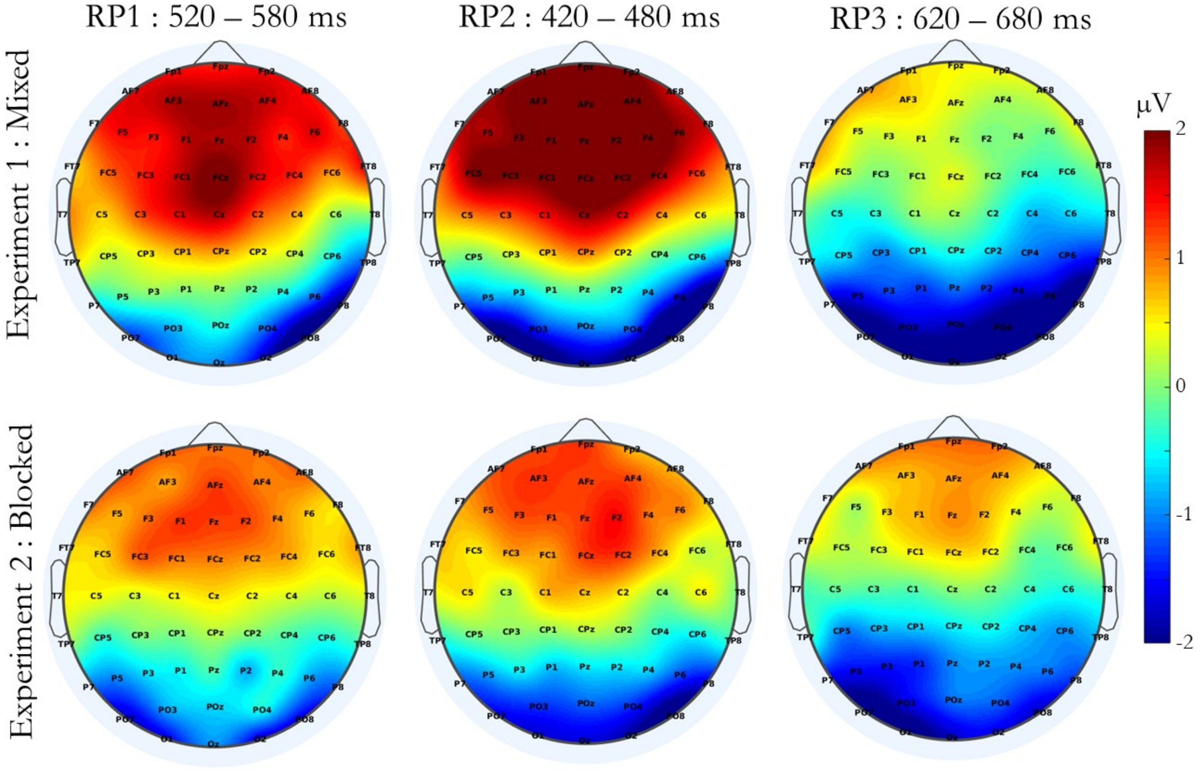

Figure 4.3. Scalp distribution of the mean activity following the presentation of the second tone. Activity during test sequence presentation was subtracted from activity from encoding sequence to eliminate low-level effects related to stimuli presentation. Measurements were taken between 20 and 80 ms post tone onset (peak at 50 ms post tone onset).

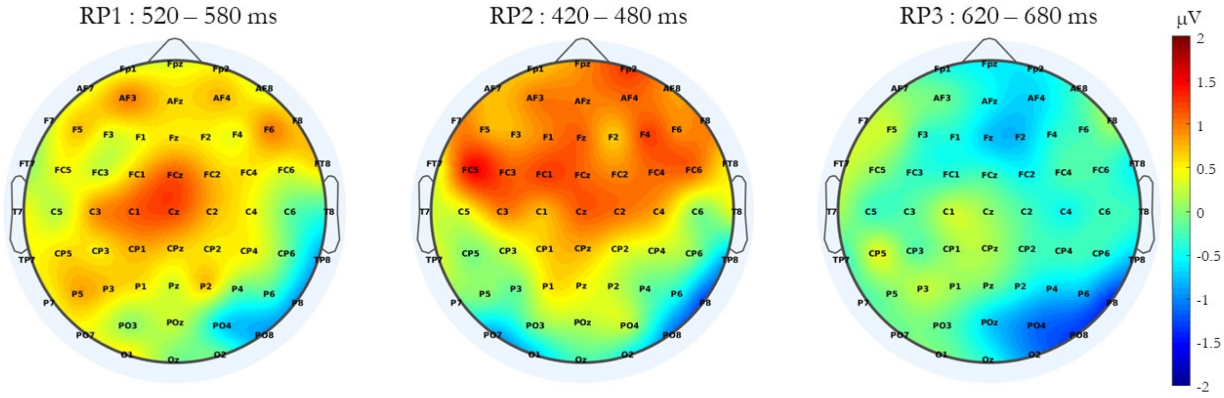

Figure 4.4. Scalp distribution of the early positivity following presentation of the second tone (average of the activity during the time window 20 to 80 ms post tone presentation). These figures were obtained by subtracting the grand average of the blocked experiment (Experiment 2) from the grand average of the mixed experiment (Experiment 1), then by subtracting the activity during test sequence from the activity in the encoding sequence. Theoretically, the activity seen in these figures reflects predictive processes as low-level effects and memory processes should have been subtracted out. Frontocentral positivity when tone was informative is clearly visible for R1 and R2, and less so for R3 (see discussion).

Differences in activity around 50 ms after presentation of stimuli of interest are shown in Figure 4.4. The positivity related to tone informativeness was maximal at frontocentral electrodes, up to frontal electrodes. Activity related to tone informativeness, obtained by computing both the difference between the two sequences (encoding and test), but also between the two experiments is shown in Figure 4.4. We measured amplitude for the average of electrodes Cz, Fcz, and Fz. Because the first tone of a tone sequence (no matter in which phase) has a specific signature, namely a larger biphasic N1-P2 response, especially since the moment of presentation was not exactly known because of the jitter preceding it, we chose to exclude it from the analysis.

The most important results of the study were obtained by subtracting ERP amplitudes observed in the test sequences from amplitudes in the encoding sequences, for each experiment, and then subtracting these differences in Experiment 2 (blocked) from the differences in Experiment 1 (mixed). This double subtraction was designed to isolate activity specifically related to predictive processes by removing low level effects caused by sensory stimulation (which was identical in encoding and testing sequences in the two experiments), as well as memory-related processes, which likely differed across encoding and testing). The resulting means are shown in Figure 4.5.

We used repeated measure ANOVAs that included a specific contrast to test the hypothesis that the second tone would be different from the third, fourth and fifth. Contrast weights were +1.5 -0.5 -0.5 -0.5 over tone position, so as to compare ERP amplitude for the second tone to mean amplitude for the third, fourth, and fifth tone.

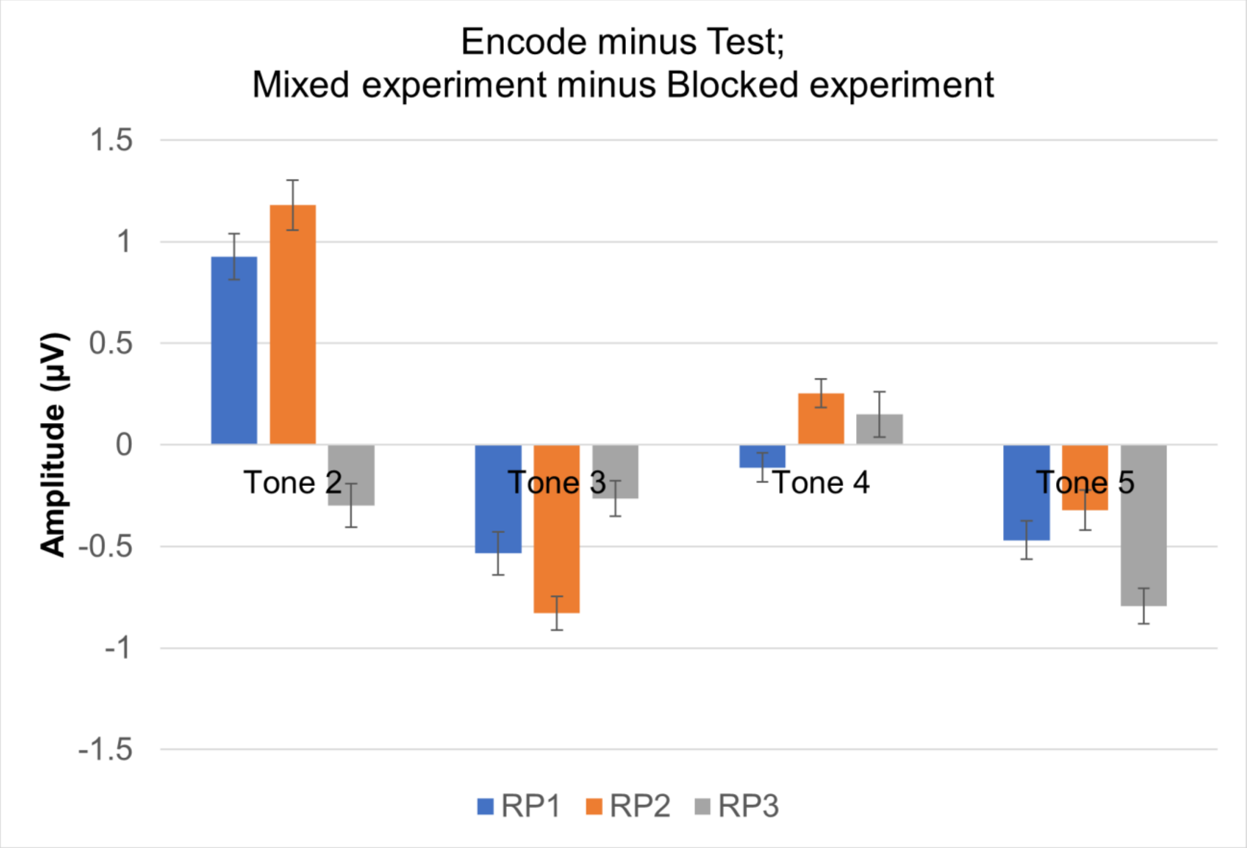

Figure 4.5. Mean amplitudes extracted from the ERP used in the statistical analysis. Amplitudes were obtained by subtracting the activity in the test sequence from the activity in the encoding sequence for each experiment, and then subtracting these differences in the blocked experiment (Experiment 2) from the differences in the mixed experiment (Experiment 1). This

subtraction isolates activity related to predictive processes by removing low level effects caused by sensory stimulation, as well as memory-related processes. Amplitude was measured from 20 ms to 80 ms following tone onset as the average of electrodes Cz, FCz, Fz. Error bars represent standard error.

For RP1 (isochronous), the repeated measure ANOVA with contrast of interest showed a significant interaction of Group x Sequence x Contrast, $F(1, 46) = 8.88$, $p = .005$, $\eta_p^2 = 0.19$, indicating there was a difference between the amplitude of the second tone and the amplitude of the three other tones that is modulated both by the between subject variable (experimental group, mixed vs blocked presentation) and by the within subject variable (sequence, encoding vs. test). For the mixed experiment, there was a significant Sequence x Contrast interaction, $F(1, 23) = 5.49$, $p = .03$, $\eta_p^2 = 0.24$. The was no Sequence x Contrast interaction for the blocked experiment, $F(1, 23) = 3.39$, $p = .08$.

There results were enough to warrant looking at factor effects for the amplitude for the second tone. The measured activity during the encoding sequence was marginally more positive in the mixed experiment ($M = -0.83$, SD = 1.81) than in the blocked experiment ($M = -0.16$, SD = 1.87), $F(1, 46) = 3.44$, $p = .07$, $\eta_p^2 = 0.07$, while they did not significantly differ for the test sequence, $F < 1$. For the other tones, all sequences, there was no difference between groups, except for Tone 5, test sequence, $F(1, 46) = 5.70$, $p = .02$, , $\eta_p^2 = 0.12$. Also, accuracy did not correlate with the difference in amplitude between encoding and test for the mixed experiment, $t(22) = -0.86$, $p = .40$, nor the blocked experiment, $t(22) = -0.37$, $p = .72$.

For RP2, we found a significant interaction of Group x Sequence x Contrast, $F(1, 46) = 7.17$, $p = .01$, $\eta_p^2 = 0.16$, as well a Sequence x Contrast interaction across groups, $F(1, 46) = 17.40$, $p < .001$, $\eta_p^2 = 0.38$. For the encoding sequence, there was a Group x Contrast interaction, $F(1, 46) = 16.66$, $p < .001$, $\eta_p^2 = 0.36$, but not for the test sequence, $F < 1$. The Sequence x Contrast interaction was significant for the mixed experiment, $F(1, 23) = 14.95$, $p < .001$, $\eta_p^2 = 0.65$, but not for the blocked experiment, $F(1, 23) = 2.59$, $p = .12$. Examination of the response to the second tone only showed an effect of Sequence across Groups, $F(1, 46) = 23.06$, $p < .001$, $\eta_p^2 = 0.50$. These effects were not created by a difference in accuracy, that did not correlate with the difference in amplitude between the encoding and test sequences for the mixed experiment, $t(22) = 0.29$, $p = .77$, nor the blocked experiment, $t(22) = 0.53$, $p = .60$.

For RP3, there was no Group x Sequence x Contrast interaction, F < 1, but there was an overall Sequence x Contrast interaction, $F(1, 46) = 6.27$, $p = .01$, $\eta_p^2 = 0.14$. The was no Group x Contrast interaction for the encoding sequence, $F < 1$, nor for the test sequence, $F(1, 46) = 1.89$, $p = .18$.

Accuracy did not correlate with the difference in amplitude between encoding and test for the mixed experiment, $t(22) = -0.36$, $p = .72$, nor the blocked experiment, $t(22) = 0.69$, $p = .50$.

4.5 Discussion

Our results show a brain signature associated with the tracking and identification of information allowing predictions of temporal structure of sequences of sounds. In Experiment 1 (mixed), we measured an increased positivity for the informative tone, that is, the second tone of the encoding sequence, only when rhythmic pattern was not known.

Both experiments were essentially the same except for the key factor that rhythmic patterns in the encoding sequence were in an uncertain context (Exp 1), or in a certain context (Exp 2). Thus, low-level factors such as the number of tones, their duration, or the need to encode tone pitches (for the memory task) cannot explain our results. Although the task did not require encoding or paying attention to the temporal patterns (only tone pitches were task-relevant), it is clear, from the comparisons across experiments, that stimuli allowing predictions of the auditory context generated a measurable electrophysiological signature.

The hypothesized effects were statistically significant for RP1 and RP2, but not for RP3. We anticipated this result by supposing the conditional probability of the onset of Tone 2 as time passed after Tone 1 would evolve dynamically. Initially, in the encoding sequence of Experiment 1, the three possible ISIs between Tone 1 and Tone 2 were equally probable (all 0.33). A tone presented at the short ISI confirmed RP1 and ruled out RP2 and RP3. If a tone was not presented at the short ISI, then the other two ISIs became equally likely (both 0.50), and the presentation of a tone at the middle ISI provided information about which rhythmic pattern was presented (confirming RP2 and ruling out RP3). If a tone was not presented at the short or middle ISI, then a tone at the long ISI became the only possible event (probability 1.0), and Tone 2, *per se*, no longer provided information about which rhythmic pattern was present in that trial. On this account, the positivity observed for Tone 2, for RP1 and RP2, but not RP3 in Figure 4.5, suggests the positivity occurs specifically when a tone provides disambiguating information about the acoustic context (in this case, about the rhythmic pattern for that trial).

The increase in positivity in the EEG signal we describe here is, to our knowledge, a new finding in the predictive coding literature. Future research will be needed to evaluate how disambiguation of hearing context affects following stimuli, or in which case it does happen. While we argue predictions and disambiguation are made pre-attentively, and do not require conscious acknowledgement, it is possible the cortical response might be modulated by general sensitivity to rhythmic pattern, musical experience, and ease of performing the main task. Although we have data on both musical expertise and accuracy in the memory task, we did not have sufficient power to detect effects of these individual difference variables. We suggest for future studies to use a passive listening task with clicks, and to test participants for their ability to discriminate subtle variations in rhythmic pattern/ISI as a more direct measure of sensitivity to rhythmic patterns.

Learning structural regularities in tone sequences could support many important functions thought to be supported by auditory scene analysis. As such, a learned structure can facilitate the processing of sounds consistent with this structure, and likely facilitate the perception of sounds that deviate from it (whether it be by pitch, timbre, or timing). According to Bregman's theory of auditory scene analysis (Bregman, 1994), the brain first generates models of alternative sound organizations, and selects the alternative that best matches the input (based on recent and distant past experience). On the basis of this theory of auditory scene analysis, the electrophysiological signature we described is likely reflecting the selection of the best/most likely sound organization. We suggest the positive ERP waveform reflects the tracking and accumulation of evidence up until selection of a specific pattern of sound organization, and updating of the internal model.

Because the temporal structure of the test sequence was always known, we expected brain activity during test to be equivalent across experiments. The activity during the second sequence was similar across experiments for R1 and R3, as predicted. However, we found a reversed pattern for RP2, compared to the encoding sequence. Isochronous sequences (RP1) showed the activity pattern we expected, which was no difference between experiments for the test sequence, when the pattern is already known. For RP2, there was a difference in amplitude, with an effect reversed from what was observed in the encoding, that is, an amplitude for the blocked experiment that was more positive than for the mixed experiment. We believe this could not be caused by predictive processes, because the pattern was already known. One possible explanation is that for the blocked experiment, a stronger grouping perception/anticipation was present, and following the strong response to the first tone, the system (neural populations) is primed and anticipatory effects were larger. More research outside the scope of this paper would be necessary to fully explain this effect.

We did not find an augmentation in P2 amplitude related with tone informativeness, or effects on the N1, in contrast with what was found by Simal et al. (2021). A possible explanation for this might be that, in the present task, the temporal aspects of sound sequences were completely irrelevant to the task. In a future experiment, it would be interesting to make the rhythmic pattern task relevant, such as by asking whether the encoding and test sequences have the same rhythmic patterns. Would bringing attention to the rhythmic patterns modulate and possibly increase the cortical response we found in the present experiments? We leave this for future research.

In conclusion, in this novel experimental design, we showed an electrophysiological response related to the tracking of informative and disambiguating information about rhythmic pattern of auditory sequences. These results suggest the brain learns regularities in the environment pre-attentively, and creates alternative models. Incoming information is then tracked and compared with these models to disambiguate the auditory scene context. These processes presumably aid perception and comprehension of rapidly changing auditory scenes.

Chapitre 5 – Discussion Générale

Les études présentées dans cette thèse ont permis de découvrir des indices électrophysiologiques liés à la résolution d'ambigüités contextuelles et à la génération de prédictions. Notamment nous avons montré un indice lié à la génération de prédiction que d'autres sons seront présentés, ainsi qu'un indice, différent, qui était lié à des prédictions d'ordre temporelles. Ces études sont cruciales en tant que preuve de faisabilité et permettront d'étudier les prédictions d'un nouveau point de vue.

5.1 Indice électrophysiologique d'un son permettant la prédiction de la présentation d'autres sons

Dans le Chapitre 2, nous avons montré que l'information que contient un son module les réponses corticales. Dans le contexte expérimental, les séquences de sons présentées ne pouvaient contenir qu'un, trois, ou cinq sons. Ainsi, si la longueur de la séquence n'était pas connue à l'avance, le deuxième et le quatrième son, s'ils étaient présentés, résolvaient une ambigüité dans le contexte et permettaient de prédire la présentation d'au moins un autre son. Ces sons informatifs étaient accompagnés de composantes de PRE N1 et P2 de plus grande amplitude que pour les autres sons. Aussi, le deuxième et le quatrième sons n'étaient pas particuliers au point de vue de leur emplacement dans la séquence ni au point de vue de leurs caractéristiques acoustiques, car les mêmes sons, présentés dans un contexte dans lequel la longueur de la séquence est connue d'avance ne généraient pas de N1 et de P2 de plus grande amplitude.

Nos résultats étaient visibles pour la N1 et la P2, mais étaient plus clairs pour la P2. La P2 n'est que très peu étudiée hors du complexe N1-P2 et il n'y a que quelques études qui se sont intéressées à sa modulation, notamment des études sur l'éveil-sommeil, le vieillissement ou encore les potentiels liés à la respiration (Crowley & Colrain, 2004). Notre étude montre un nouvel exemple de modulation de la P2 au moins en partie indépendante de la modulation de la N1. Alors que la N1 est sensible à une myriade de facteurs tels que l'attention, la répétition, l'intensité des sons et l'amorçage (Näätänen & Picton, 1987), ce qui ajoute du bruit, la modulation de la P2 que nous avons observée semble être spécifiquement liée aux processus prédictifs.

Dans l'expérience, le nombre de sons présentés n'était pas directement utile à la tâche de mémoire qui portait sur un jugement concernant la hauteur des sons. Nous avons toutefois considéré la possibilité que la modulation de la P2 soit une P3a, visible, comme la P2, aux électrodes frontocentrales, avec une latence similaire. Une P3a aurait pu être générée par un mécanisme de changement de tâche « *task switching* » par la mise à jour de l'empan mnésique nécessaire à l'exécution de la tâche (Barcelo et al., 2006). Toutefois, l'analyse de décomposition des signaux que nous avons effectué a mis en lumière des activations modulées par les processus prédictifs aux électrodes temporales, qui ne peuvent pas être expliquées par une P3a.

Cette étude nous a donc permis d'observer les réponses cérébrales lorsqu'un son apporte de l'information qui permet d'effectuer des prédictions. Ce processus semble se faire de manière inconscience, ou tout du moins, sans lien avec les buts de la tâche. Il s'agit également d'un processus rapide, l'activité liée aux prédictions étant observable dès 100 ms suivant la présentation du son informatif.

5.2 Oscillations liées à un son permettant la prédiction de la présentation d'autres sons

Dans le Chapitre trois de cette thèse, nous souhaitions mieux comprendre les résultats du Chapitre deux, c'est à dire d'élucider de quels processus prédictifs en particulier peuvent provenir les modulations d'amplitude de la P2. Pour cela l'étude des oscillations est particulièrement utile, notamment car elles permettent d'avoir une idée sur le type de communication intracérébrale qui s'effectue (Başar et al., 2001; Fries, 2015). Nous avons montré que les sons informatifs, qui permettent de prédire qu'au moins un autre son sera présenté, sont accompagnés d'une augmentation d'amplitude thêta (4 – 7 Hz). Cette augmentation était phasique et maximale aux électrodes frontocentrales. Dans la littérature, l'activité dans la bande de fréquence thêta est liée avec des changements d'excitabilité neuronale, ainsi qu'avec des changements attentionnels (Fries, 2015). Il est possible que l'augmentation thêta, qui semble entraîner les modulations de l'activité de PRE, reflète un ajustement de l'excitabilité des neurones pour se préparer à traiter le stimulus dont la présentation vient d'être prédite. Cette idée est également cohérente avec les études qui impliquent thêta dans le recrutement d'aires supérieures afin d'implémenter un contrôle cognitif approprié lorsque le contexte est incertain (Cavanagh & Frank, 2014). Cette modulation claire de thêta est une preuve supplémentaire en faveur de notre interprétation de l'augmentation de l'amplitude de la P2 comme étant un signe de prédiction.

5.3 Indice électrophysiologie lié à l'identification du contexte temporel

Dans l'étude du Chapitre 4, nous souhaitions tester un autre aspect des prédictions dans le domaine auditif. Alors que les deux premières études peuvent être décrites comme des études de prédiction d'occurrence, l'étude de ce chapitre s'intéresse aux prédictions d'ordre temporelle. Puisque les études présentées dans cette thèse sont nouvelles et n'ont jamais été faites, il était raisonnable de se poser la question à savoir si les indices électrophysiologiques liés à la génération de prédictions temporelles seraient également visibles dans des modulations de la N1 et de la P2. Contrairement à nos hypothèses à priori, nous n'avons pas trouvé de modulation de la N1 et de la P2. À la place, nous avons trouvé que lorsqu'un son résout l'ambigüité contextuelle et permet d'identifier le patron rhythmique entendu, l'activité de PRE devient plus positive que pour le même son dans un contexte certain avec une latence extrêmement courte d'environ 50 ms. Cette modulation de positivité précoce, visible aux électrodes frontales et frontocentrales, n'a, à notre connaissance, jamais été décrite. Elle était claire pour le patron rhythmique isochrone RP1, ainsi que le patron durant lequel le deuxième son est présenté après un cours intervalle (RP2). Tel que discuté dans le chapitre, lorsque l'ISI entre le premier son et le son informatif (le second) est long, le patron rhythmique peut être déduit sans avoir besoin d'avoir entendu le son.

En ce qui concerne l'absence de modulation de la P2, il est possible que dans l'étude du Chapitre 2, même si la connaissance de la longueur des séquences n'est pas directement utile à la tâche, qu'il y a un avantage à avoir cette information pour encoder l'identité des sons. Ainsi, il est possible que la modulation sur la P2 soit liée au recrutement d'aires supérieures pour ajuster le contrôle cognitif, ce qui n'est pas nécessaire dans l'expérience du Chapitre 4 dans laquelle le patron temporel est réellement indépendant de la tâche. Ces hypothèses seront à tester dans une étude future.

5.4 Expériences futures

Une des limites des études présentées dans les Chapitre 2 et 3 de cette thèse, est que nous n'avons pas d'information sur la localisation des sources qui répondent aux sont qui apportent de l'information permettant de faire des prédictions. L'expérience pour tester ceci, en utilisant la magnétoencéphalographie, ainsi qu'une image structure du cerveau de chaque participant est prête, mais n'a pas pu être faite dans le contexte actuel. Cette étude permettra de déterminer avec plus de précision les sources corticales qui réponse à un son contenant de l'information sur le contexte, ainsi que d'obtenir une idée de la connectivité reflétée dans les oscillations thêta.

Maintenant que nous savons que les effets liés aux prédictions sont observables même si les participants n'en ont pas conscience, on peut imaginer que des paradigmes simplifiés dans lesquels les stimuli sont présentés passivement pourraient permettre d'observer les processus prédictifs sans avoir à contrôler pour des effets liés à la tâche. De plus afin de simplifier la tâche pour éviter le plus de facteurs confondants, nous pouvons utiliser des clics plutôt que des sons purs afin de ne pas risquer de créer d'interaction entre la « mélodie » artificielle et le patron rhythmique ou la longueur des séquences. Une telle expérience serait également plus adaptée pour tester des populations pour qui faire une tâche qui requiert de la concentration pendant 2 heures n'est pas toujours facile ni faisable. Ainsi une étude d'écoute passive permettrait de tester les enfants, les personnes âgées, ou encore des personnes souffrant de troubles neuropsychologiques.

Pour la troisième étude présentée dans le Chapitre 4, l'expérience suivante sera de rendre le patron rhythmique pertinent à la tâche, par exemple demandant aux participants de juger si deux séquences qui contiennent le même nombre de sons (sons de hauteur identique entre les séquences) ont le même patron temporel. Cette tâche qui devrait être facile, ne semble pas l'être selon des données préliminaires puisque les taux de bonne réponse variaient entre 68 et 99 % environ (après avoir retiré deux participants qui avaient moins de 60% de bonnes réponses). Nous pourrions également remplacer les sons purs pour des clics afin de voir si cela a un effet sur les taux de succès. Il est également possible que la capacité à discriminer les intervalles silencieux est très variable de personne à personne et il s'agit d'une variable qui devra être contrôlée.

Afin de prouver que l'activité cérébrale que nous avons décrit est liée à des processus prédictifs, des études subséquentes pourraient s'intéresser à répliquer les effets décrits dans les études de cette thèse en testant également si un stimulus déviant/inattendu par rapport à la prédiction génère bien un signal d'erreur de prédiction

5.5 Conclusion

En conclusion, les expériences présentées dans cette thèse sont novatrices et posent une base solide pour le développement d'études subséquentes visant à étudier les processus cérébraux prédictifs. Dans le domaine auditif les prédictions semblent avoir une importance particulière car les sons sont de nature fugace et le déroulement dans le temps rend crucial le fit de traiter efficacement les stimuli entrants. Nous avons montré que des sons qui permettent de prédire la présentation d'autres sons ou de prédire quand d'autres sons seront présentés sont accompagnés d'une signature électrophysiologique visible dans les PRE, ainsi que dans l'activité oscillatoire sous-jacente. Nous avons montré que les prédictions auditives se font de manière rapide, dynamique, pré-attentive et sans nécessiter une intention consciente de faire ces prédictions.

Bibliographie

Alunni-Menichini, K., Guimond, S., Bermudez, P., Nolden, S., Lefebvre, C., & Jolicoeur, P. (2014). Saturation of auditory short-term memory causes a plateau in the sustained anterior negativity event-related potential. *Brain Research*, *1592*, 55–64. https://doi.org/10.1016/j.brainres.2014.09.047

Barcelo, F., Escera, C., Corral, M. J., & Periáñez, J. A. (2006). Task switching and novelty processing activate a common neural network for cognitive control. *Journal of Cognitive Neuroscience*, *18*(10), 1734–1748. https://doi.org/10.1162/jocn.2006.18.10.1734

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *39*(2–3), 241–248. https://doi.org/10.1016/s0167-8760(00)00145-8

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N. J., & Schröger, E. (2008). Rapid extraction of auditory feature contingencies. *NeuroImage*, *41*(3), 1111–1119. https://doi.org/10.1016/j.neuroimage.2008.03.040

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *The Journal of Neuroscience: The*

Official Journal of the Society for Neuroscience, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Berker, E. A., Berker, A. H., & Smith, A. (1986). Translation of Broca's 1865 report. Localization of speech in the third left frontal convolution. *Archives of Neurology*, *43*(10), 1065–1072. https://doi.org/10.1001/archneur.1986.00520100069017

Bobes, M. A., Quiñonez, I., Perez, J., Leon, I., & Valdés-Sosa, M. (2007). Brain potentials reflect access to visual and emotional memories for faces. *Biological Psychology*, *75*(2), 146–153. https://doi.org/10.1016/j.biopsycho.2007.01.006

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, *137*(2), 321–332. https://doi.org/10.1016/j.jneumeth.2004.03.002

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between "what" and "when" in the auditory system: Temporal predictability enhances repetition suppression. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

D’agostino Sr, R. B., & Russell, H. K. (2005). Scree test. *Encyclopedia of Biostatistics*, *7*.

Danielmeier, C., & Ullsperger, M. (2011). Post-Error Adjustments. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00233

Dell’Acqua, R., Doro, M., Dux, P. E., Losier, T., & Jolicœur, P. (2016). Enhanced frontal activation underlies sparing from the attentional blink: Evidence from human electrophysiology. *Psychophysiology*, *53*(5), 623–633. https://doi.org/10.1111/psyp.12618

Dell’Acqua, R., Dux, P. E., Wyble, B., Doro, M., Sessa, P., Meconi, F., & Jolicœur, P. (2015). The attentional blink impairs detection and delays encoding of visual information: Evidence from human electrophysiology. *Journal of Cognitive Neuroscience*, *27*(4), 720–735. https://doi.org/10.1162/jocn_a_00752

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: Challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Donchin, E. (1981). Presidential address, 1980. Surprise!...Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Drisdelle, B. L., Aubin, S., & Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory: ICA correction versus epoch rejection. *Psychophysiology*, *54*(1), 83–99. https://doi.org/10.1111/psyp.12675

Fava, J. L., & Velicer, W. F. (1992). The Effects of Overextraction on Factor and Component Analysis. *Multivariate Behavioral Research*, *27*(3), 387–415. https://doi.org/10.1207/s15327906mbr2703_5

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement*, *56*(6), 907–929. https://doi.org/10.1177/0013164496056006001

Folyi, T., Fehér, B., & Horváth, J. (2012). Stimulus-focused attention speeds up auditory processing. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *84*(2), 155–163. https://doi.org/10.1016/j.ijpsycho.2012.02.001

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, *75*, 27–35. https://doi.org/10.1016/j.neuroimage.2013.02.044

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/s1364-6613(00)01839-8

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Fuentemilla, Ll., Marco-Pallarés, J., Münte, T. F., & Grau, C. (2008). Theta EEG oscillatory activity and auditory change detection. *Brain Research*, *1220*, 93–101. https://doi.org/10.1016/j.brainres.2007.07.079

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*(6), 627–640. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., Zatorre, R., Robitaille, N., & Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *NeuroImage*, *94*, 96–106. https://doi.org/10.1016/j.neuroimage.2014.03.020

Grimm, S., Escera, C., Slabu, L., & Costa-Faidella, J. (2011). Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. *Psychophysiology*, *48*(3), 377–384. https://doi.org/10.1111/j.1469-8986.2010.01073.x

Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory. *Psychophysiology*, *48*(11), 1500–1509. https://doi.org/10.1111/j.1469-8986.2011.01234.x

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. *The Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. https://doi.org/10.1162/089892999563328

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. https://doi.org/10.1016/j.biopsycho.2008.04.001

Huberle, E., & Karnath, H.-O. (2012). The role of temporo-parietal junction (TPJ) in global Gestalt perception. *Brain Structure and Function*, *217*(3), 735–746. https://doi.org/10.1007/s00429-011-0369-y

Hulse, S. H., Takeuchi, A. H., & Braaten, R. F. (1992). Perceptual Invariances in the Comparative Psychology of Music. *Music Perception: An Interdisciplinary Journal*, *10*(2), 151–184. https://doi.org/10.2307/40285605

Jung, T.-P., Makeig, S., McKeown, M. J., Bell, A. J., Lee, T.-W., & Sejnowski, T. J. (2001). Imaging brain dynamics using independent component analysis. *Proceedings of the IEEE*, *89*(7), 1107–1122.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Koelsch, S., Gunter, T. C., Schröger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *Neuroreport*, *12*(7), 1385–1389. https://doi.org/10.1097/00001756-200105250-00019

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: “nonmusicians” are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541. https://doi.org/10.1162/089892900562183

Krause, C. M., Heikki Lang, A., Laine, M., Kuusisto, M., & Pörn, B. (1996). Event-related. EEG desynchronization and synchronization during an auditory memory task. *Electroencephalography and Clinical Neurophysiology*, *98*(4), 319–326. https://doi.org/10.1016/0013-4694(96)00283-0

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition*, *69*(1), 127–137. https://doi.org/10.1016/j.bandc.2008.06.004

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, *40*(5), 806–817. https://doi.org/10.1111/1469-8986.00081

Lawrance, E. L. A., Harper, N. S., Cooke, J. E., & Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *The Journal of the Acoustical Society of America*, *135*(6), EL357-363. https://doi.org/10.1121/1.4879667

Lefebvre, C., & Jolicœur, P. (2016). Memory for pure tone sequences without contour. *Brain Research*, *1640*(Pt B), 222–231. https://doi.org/10.1016/j.brainres.2016.02.025

Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., Zatorre, R. J., & Jolicœur, P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia*, *51*(13), 2939–2952. https://doi.org/10.1016/j.neuropsychologia.2013.08.003

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford university press.

Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209.

May, P. J. C., & Tiitinen, H. (2004). The MMN is a derivative of the auditory N100 response. *Neurology & Clinical Neurophysiology: NCN*, *2004*, 20.

May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

May, P., & Tiitinen, H. (2001). Human cortical processing of auditory events over time. *Neuroreport*, *12*(3), 573–577. https://doi.org/10.1097/00001756-200103050-00028

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(5), 545–551. https://doi.org/10.1093/cercor/bhh155

Mondor, T. A. (1999). Predictability of the cue-target relation and the time-course of auditory inhibition of return. *Perception & Psychophysics*, *61*(8), 1501–1509. https://doi.org/10.3758/bf03213113

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233. https://doi.org/10.1017/S0140525X00078407

Näätänen, R., & Gaillard, A. W. K. (1983). 5 The Orienting Reflex and the N2 Deflection of the Event-Related Potential (ERP). In A. W. K. Gaillard & W. Ritter (Eds.), *Advances in Psychology* (Vol. 10, pp. 119–141). North-Holland. https://doi.org/10.1016/S0166-4115(08)62036-1

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., & Kreegipuu, K. (2012). The mismatch negativity (MMN). In *The Oxford handbook of event-related potential components* (pp. 143–157). Oxford University Press.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*(1), 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, *14*(1–2), 53–98. https://doi.org/10.1016/0301-0511(82)90017-5

Neszmélyi, B., & Horváth, J. (2017). Consequences matter: Self-induced tones are used as feedback to optimize tone-eliciting actions. *Psychophysiology*, *54*, 904–915. https://doi.org/10.1111/psyp.12845

Nolden, S., Bermudez, P., Alunni-Menichini, K., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2013). Electrophysiological correlates of the retention of tones differing in timbre in auditory short-term memory. *Neuropsychologia*, *51*(13), 2740–2746. https://doi.org/10.1016/j.neuropsychologia.2013.09.010

Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., & Jolicoeur, P. (2013). The retention of simultaneous tones in auditory short-term memory: A magnetoencephalography study. *NeuroImage*, *82*, 384–392. https://doi.org/10.1016/j.neuroimage.2013.06.002

Notter, M. P., Hanke, M., Murray, M. M., & Geiser, E. (2019). Encoding of Auditory Temporal Gestalt in the Human Brain. *Cerebral Cortex*, *29*(2), 475–484. https://doi.org/10.1093/cercor/bhx328

Palva, S., & Palva, J. M. (2007). New vistas for alpha-frequency band oscillations. *Trends in Neurosciences*, *30*(4), 150–158. https://doi.org/10.1016/j.tins.2007.02.001

Pantazis, D., Simpson, G. V., Weber, D. L., Dale, C. L., Nichols, T. E., & Leahy, R. M. (2009). A novel ANCOVA design for analysis of MEG data with application to a visual attention study. *NeuroImage*, *44*(1), 164–174. https://doi.org/10.1016/j.neuroimage.2008.07.012

Pesonen, M., Björnberg, C. H., Hämäläinen, H., & Krause, C. M. (2006). Brain oscillatory 1–30Hz EEG ERD/ERS responses during the different stages of an auditory memory search task. *Neuroscience Letters*, *399*(1), 45–50. https://doi.org/10.1016/j.neulet.2006.01.053

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. I. Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, *36*(2), 179–190. https://doi.org/10.1016/0013-4694(74)90155-2

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology. General*, *134*(1), 73–92. https://doi.org/10.1037/0096-3445.134.1.73

Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Research*, *745*(1), 134–143. https://doi.org/10.1016/S0006-8993(96)01140-7

Rosburg, T., Trautner, P., Dietl, T., Korzyukov, O. A., Boutros, N. N., Schaller, C., Elger, C. E., & Kurthen, M. (2005). Subdural recordings of the mismatch negativity (MMN) in patients with focal epilepsy. *Brain: A Journal of Neurology*, *128*(Pt 4), 819–828. https://doi.org/10.1093/brain/awh442

Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A., & Binder, J. R. (2006). Attentional Modulation in the Detection of Irrelevant Deviance: A Simultaneous ERP/fMRI Study. *Journal of Cognitive Neuroscience*, *18*(5), 689–700. https://doi.org/10.1162/jocn.2006.18.5.689

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006). Relevance of EEG alpha and theta oscillations during task switching. *Experimental Brain Research*, *170*(3), 295–301. https://doi.org/10.1007/s00221-005-0211-y

Schadow, J., Lenz, D., Dettler, N., Fründ, I., & Herrmann, C. S. (2009). Early gamma-band responses reflect anticipatory top-down modulation in the auditory cortex. *NeuroImage*, *47*(2), 651–658. https://doi.org/10.1016/j.neuroimage.2009.04.074

Shalgi, S., & Deouell, L. Y. (2007). Direct evidence for differential roles of temporal and frontal components of auditory change detection. *Neuropsychologia*, *45*(8), 1878–1888. https://doi.org/10.1016/j.neuropsychologia.2006.11.023

Sharbrough, F., Chatrian, G.-E., Lesser, R. P., Luders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*(2), 200–202.

Simal, A., Bermudez, P., Lefebvre, C., Vachon, F., & Jolicœur, P. (2021). Signal informativeness for sequence structure modulates human auditory cortical responses. *Psychophysiology*, *58*(3), e13745. https://doi.org/10.1111/psyp.13745

Simal, A., & Jolicoeur, P. (2020). Scanning acoustic short-term memory: Evidence for two subsystems with different time-course and memory strength. *International Journal of Psychophysiology*, *155*, 105–117. https://doi.org/10.1016/j.ijpsycho.2020.06.004

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, *13*(3), 393–405. https://doi.org/10.1016/S0926-6410(01)00131-8

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637. https://doi.org/10.3758/BF03196525

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In *The Oxford handbook of event-related potential components* (pp. 397–439). Oxford University Press.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, 879716. https://doi.org/10.1155/2011/879716

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Tallon-Baudry, C., Bertrand, O., Wienbruch, C., Ross, B., & Pantev, C. (1997). Combined EEG and MEG recordings of visual 40 Hz responses to illusory triangles in human. *NeuroReport*, *8*(5), 1103.

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Tsogli, V., Jentschke, S., Daikoku, T., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-42066-4

Turk-Browne, N. B. (2012). Statistical Learning in Perception. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3182–3185). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_1707

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *47A*(3), 631–650. https://doi.org/10.1080/14640749408401131

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163. https://doi.org/10.1027/0269-8803.21.34.147

Winkler, I., Denham, S., & Escera, C. (2013). Auditory Event-related Potentials. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational Neuroscience* (pp. 1–29). Springer New York. https://doi.org/10.1007/978-1-4614-7320-6_99-1

Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, *39*(6), 609–620. https://doi.org/10.1016/0013-4694(75)90073-5

Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology. Supplement*, *44*, 102–109.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*(5), 615–619. https://doi.org/10.1017/s0048577298000183

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7