Stage de recherche
Master ATIAM
Acoustique Traitement de signal et Informatique
Appliqués à la Musique
Année universitaire 2016-2017

Contrôle de sa propre voix pendant la production de discours : une étude EEG

Lou Seropian

Encadrement : Jean-Julien Aucouturier & Laura Rachman
(Equipe Perception et Design Sonores de l'IRCAM)

19 Août 2017

Résumé

Lorsque nous parlons, nous ajustons inconsciemment notre voix afin de produire les sons attendus, à savoir ceux qui pourront être interprétés par notre interlocuteur. Plusieurs études électrophysiologiques autour de ce contrôle automatique de notre propre voix ont été menées ces dernières années et une théorie a été émise sur les changements sensoriels causés par nos propres actions. Cette théorie est basée sur des modèles dits de prédiction interne selon lesquels chaque représentation d'une commande motrice est traduite par un événement sensoriel au niveau cérébral. Dans la parole, le seul élément qui nous permet de contrôler les sons que nous produisons est le retour auditif que nous entendons en même temps que nous parlons. Ce retour auditif joue donc un rôle crucial au sein de ce modèle de prédiction interne. Plusieurs phénomènes neuro-physiologiques en réponse à la production de discours ont été découverts. Le premier phénomène observé est une suppression du composant évoqué N1 directement induit par le fait que la parole est produite par soi. Le deuxième phénomène s'observe plus tardivement et concerne le composant P2. En effet celui-ci est significativement plus grand lorsqu'une erreur est détectée dans le retour auditif perçu par rapport au retour auditif attendu lors de la production de sons. Jusqu'ici ces résultats ont été montrés principalement sur la production de phonèmes isolés mais, pour des raisons pratiques, pas pour de la parole continue. L'objectif de ce stage est de répliquer ces résultats pour la production de phrases complètes. Pour ce faire, nous avons modifié un outil de transformation vocal temps-réel (DAVID) pré-existant dans l'équipe pour l'adapter à un paradigme de rétroaction vocale modifiée avec captation EEG. Avec ce nouvel outil, deux expériences ont été mises en place : une expérience comportementale permettant de mesurer les seuils de détection de changement de hauteur des participants dans des conditions de production et d'écoute et une expérience électro-physiologique avec laquelle nous avons pu récolter des données EEG pour en extraire les potentiels évoqués (PE, en anglais ERPs) dans ces deux même conditions. Les données comportementales ne permettent pas de conclure à une différence significative du seuil de détection d'erreur (environ +50 cent) entre production et perception. Les données EEG confirment en partie les idées du modèle de prédiction interne : nous observons pour la condition production un composant N1 significativement plus petit que pour la condition écoute. D'autre part, nous montrons que le paradigme de parole continue rend difficile l'analyse traditionnelle des composants tardifs, et nous proposons pour ce faire plusieurs pistes d'analyse. Globalement, ce travail fournit un nouvel outil logiciel (bientôt rendu disponible sur la plateforme IRCAM forum.net) et démontre la faisabilité d'études EEG de la production de parole dans un contexte réaliste de discours continu. Ce paradigme portant pour l'instant sur la propre voix du participant pourra par la suite être étendu à l'analyse des mécanismes de traitement de la voix d'un tiers et ainsi permettre une comparaison des deux types de traitement. Une dimension musicale pourra également être explorée en confrontant voix parlée et voix chantée.

Mots-clés : Voix parlée, discours continu, auto-contrôle, prédiction interne, retour auditif modifié, EEG

Abstract

As we talk, we unconsciously adjust our speech to ensure it sounds the way we intend it to, and thus make ourselves understood. The electro-physiological indices of such vocal control mechanisms are a well-researched domain of study, summarized by the "internal forward model" theory : every motor command involved in speech production is translated into a predicted sensory event (efference copy), that can be compared to the actual event when it is perceived, e.g. via auditory feedback. Several neurophysiological phenomena are believed to correspond to this mechanism. The first one, speaking induced suppression (SIS), is the event-related suppression of the N1 ERP component directly linked to a self-made action. The second one is observed in later components like P2, which amplitude is significantly larger in a production than in passive listening when an error in the feedback is detected (speech perturbation response enhancement, or SPRE). So far, for practical reasons, these phenomena have only been shown while vocalizing isolated vowels. The goal of this internship is to replicate these findings on more complex utterances (e.g. whole sentences) and to extend existing altered auditory feedback paradigms to the case of continuous speech. To do so, we extended the functionalities of an existing voice transformation tool (DAVID) to adapt it to altered auditory feedback paradigms, and make it suitable to EEG experiments. Using this tool, two different experiments were conducted : first, a behavioral experiment which allowed us to compare participants pitch change detection thresholds in active (speaking) and passive (listening) conditions. Second, a EEG/ERP experiment also comparing the same conditions. Behavioral results did not show any significant difference between production and passive listening thresholds (comparable at +50 cents). ERP results partially supported the internal forward model theory : a significant suppression of the N1 component was found for the production condition ; in addition, we showed that continuous speech makes the traditional analysis of P2 components difficult, and propose several alternative strategies for analysis. On the whole, this work provides a novel experimental tool (to be made available on the IRCAM Forum.net plateform) and demonstrates the feasibility of studying speech production ERPs in more realistic continuous speech contexts. While the present work only focuses on the sound of a participant's own voice, we plan to extend it to study the processing of others' voices, as well as to compare error monitoring in spoken versus singing voice production.

Keywords : Running speech, self-monitoring, internal forward model, altered auditory feeback, EEG

Remerciements

Je voudrais d'abord remercier tout particulièrement mes encadrants Jean-Julien Aucouturier et Laura Rachman pour toutes les connaissances qu'ils ont pu m'apporter dans le domaine des neurosciences et pour tout le temps qu'ils m'ont consacré au cours de ce stage. Je leur suis également reconnaissante de m'avoir fait confiance dès le début de notre collaboration et remercie leur soutien tout au long de l'étude. J'aimerais ensuite remercier le reste de l'équipe pour leur accueil chaleureux et leurs précieux conseils, avec un remerciement particulier à Pablo pour ses talents de débogueur.

Je remercie également toute l'équipe pédagogique du master ATIAM pour le bagage scientifique important qui nous a été délivré cette année. Enfin, je remercie évidemment tous mes camarades du master ATIAM pour leur bonne humeur au quotidien et le soutien apporté par chacun au cours de ce premier semestre mouvementé.

Table des matières

1 **Indices EEG du contrôle de la voix** **9**
- 1.1 Paradigme du contrôle de la voix . . . 9
- 1.2 Introduction à l'EEG et aux ERP . . . 9
- 1.3 Indices relatifs au contrôle de la voix . . . 10
- 1.4 Hypothèses . . . 12

2 **Expérimentation** **14**
- 2.1 Protocole expérimental . . . 14
  - 2.1.1 Tâche comportementale . . . 14
  - 2.1.2 Tâche EEG . . . 15
- 2.2 Acquisition des données . . . 16
- 2.3 Traitement des signaux récoltés . . . 18
- 2.4 Analyses statistiques utilisées . . . 20

3 **Résultats** **21**
- 3.1 Données comportementales . . . 21
- 3.2 Données EEG . . . 22
  - 3.2.1 Composant N1 . . . 22
  - 3.2.2 Composants P1, P2 . . . 22

4 **Discussion** **24**

Introduction

Présentation de l'organisme d'accueil

Fondé par Pierre Boulez en 1977, l'Institut de Recherche et Coordination Acoustique Musique (IRCAM) associé au Centre Pompidou est aujourd'hui l'un des plus grands centres de recherche publique au monde se consacrant à la création musicale et à la recherche scientifique. Les travaux qui y sont menés sont le fruit de la recherche de plusieurs équipes : Espaces Acoustiques et Cognitifs, Perception et Design Sonores, Analyse et Synthèse des Sons, Systèmes et Signaux Sonores : Audio/Acoustique, instruMents, Représentations Musicales, Analyse des Pratiques Musicales et Interaction Son Musique Mouvement. L'IRCAM a pour mission fondamentale de susciter une interaction féconde entre recherche scientifique, développement technologique et création musicale contemporaine. Cette articulation constitue, depuis sa fondation il y a quarante ans, le principal axe structurant l'ensemble de ses activités. L'un des enjeux majeurs est de contribuer, par les apports des sciences et techniques, au renouvellement de la création musicale contemporaine, et de l'expression artistique en général.

Le projet de l'équipe *Perception et Design Sonores*, au sein de laquelle le stage a été effectué, porte sur la perception et la cognition des sons en combinant des connaissances en pyschoacoustique, en traitement/synthèse du signal, en psychologie et en neurosciences cognitives. Les travaux concernent aussi bien la caractérisation perceptive des sons que les mécanismes cognitifs mis en jeu pour les identifier. Initialement basée sur les sons environnementaux, l'équipe a étendu l'objet de ses recherches aux imitations vocales, aux sons musicaux et à la voix afin de mieux comprendre les processus cognitifs mis en jeu, d'une part, par l'identification d'une source sonore, et d'autre part, par le traitement émotionnel des sons. Les applications des ces travaux concernent d'une part le design sonore (conception de règles, d'outils, de méthodes) visant à faciliter la création de sons fonctionnels et agréable pour notre environnement urbain et, dans une moindre mesure, les applications cliniques notamment dans le domaine de la santé mentale. Ce stage s'inscrit plus particulièrement dans le projet CREAM (*Cracking the Emotional Code of Music*, ERC StG 2014-2019) qui se donne pour objectif de produire les technologies et les connaissances permettant de caractériser comment les signaux sonores (musique ou parole) activent les mécanismes cérébraux émotionnels[1].

Le travail de ce stage est axé sur les neurosciences et la perception de sa propre voix. Nous y avons mené une étude d'électro-encéphalographie (EEG), impliquant le développement d'un dispositif expérimental couplant technologie audio et captation EEG, la conception de l'expérience, la collecte de données auprès de N=21 participants expérimentaux, le traitement informatique des signaux récoltés et l'analyse statistique des indicateurs extraits de ces derniers.

Précédents travaux effectués dans l'équipe

La problématique de la perception de sa propre voix est explorée depuis plusieurs années dans le cadre du projet ERC CREAM. L'équipe a tout d'abord créé (en collaboration avec Petter Johansson et Lars Hall de l'Université de Lund en Suède, Katsumi Watanabe de l'Université Waseda au Japon) un algorithme de traitement du signal (appelé DAVID - nous y reviendrons) qui permet de modifier subrepticement le ton émotionnel de la voix d'un participant, en temps-réel, au fur et à mesure qu'il ou elle parle, pour la faire paraître plus joyeuse, triste ou anxieuse (Figure 1). Pour des auditeurs extérieurs, les transformations sont émotionnellement claires et naturelles, pourtant les locuteurs dont la voix est ainsi modifiée ne détectent pas la manipulation. De plus, dû au fait de s'écouter parler avec un ton de voix modifié, l'équipe a montré que l'état émotionnel des participants change dans la direction de l'émotion manipulée : en s'entendant parler de façon plus joyeuse, les participants deviennent plus joyeux eux-mêmes (Figure 2). Cet effet, dit de rétroaction vocale émotionnelle, a été publié dans la revue PNAS en Janvier 2016[2].

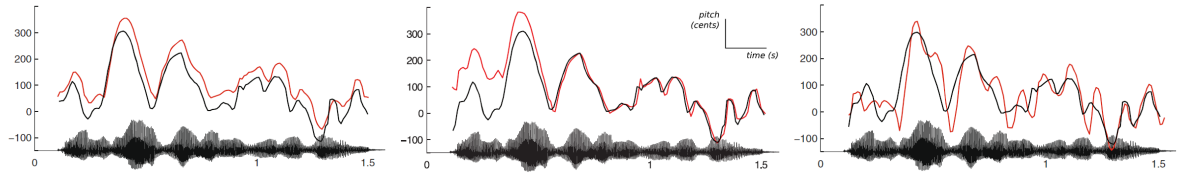

FIGURE 1 – Trois des transformations de *pitch*[1] réalisées par l'outil logiciel DAVID, pré-existant au stage : (1) une élévation ou diminution statique du *pitch* (*"pitch shifting"*), paramétrée en amplitude (ex. + 50 cents), (2) une modification dynamique du *pitch* au début des uttérances paramétrée avec des tronçons d'exponentielle et une durée (ex. 500ms), et déclenchée après un passage au silence (*"inflection"*), (3) une modulation périodique du *pitch*, paramétrée en fréquence (ex. 8Hz) et en profondeur (ex. 50 cents) ("vibrato"). Noir = *pitch* original, rouge = *pitch* modifié. Les manipulations sont implémentées de façon à être utilisable en temps-réel, avec moins de 20ms de latence. Figure adaptée de [3].

Un des points marquants de ce premier résultat est le fait que les participants, malgré l'effet de rétroaction sur leurs propres émotions, ne semblent pas détecter la manipulation de façon consciente, alors que les changements sont évidents pour des auditeurs extérieurs. Il semble donc y avoir des différences implicites entre le fait de traiter la prosodie émotionnelle sur sa propre voix et celle de tiers.

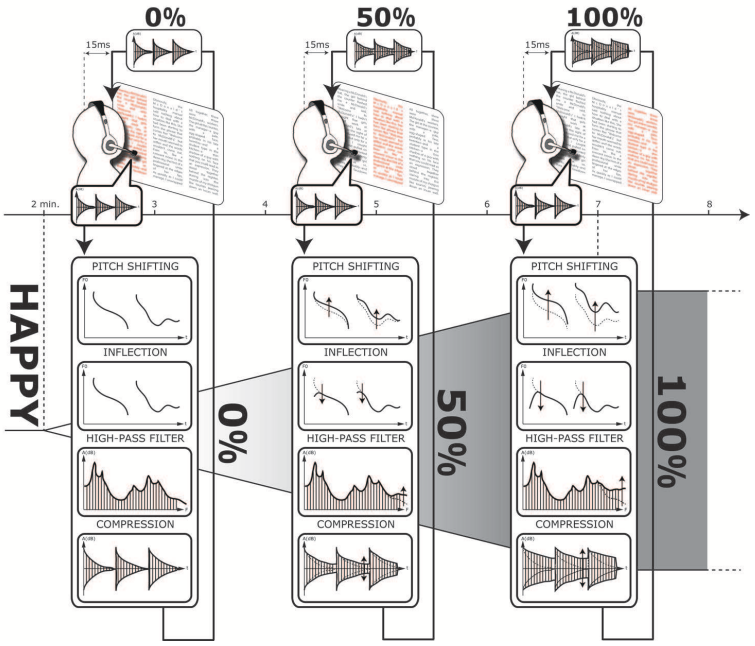

FIGURE 2 – Dispositif expérimental de rétroaction vocale émotionnelle, pré-existant au stage : les participants lisent un texte à voix haute, tout en s'entendant au casque. A leur insu, l'expérimentateur modifie avec des algorithmes de traitement du signal le retour sonore de leur propre voix pour la rendre progressivement plus expressive. Figure adaptée de [2].

Afin d'étudier ces différences, l'équipe a ensuite mené une étude EEG utilisant le paradigme expérimental de la négativité de déviance (*mismatch negativity*), une composante du potentiel cérébral évoqué (PE, en anglais *event-related potential* - *ERP* ; voir section 1.2) par un stimulus déviant inséré

1. La hauteur (*pitch* en anglais) est la qualité de la sensation qui permet d'évaluer si un son est grave ou aigu. Elle est approximativement le correspondant perceptif de la fréquence[4]. On utilisera ici les cents pour qualifier la différence de hauteur, la valeur en cents de l'intervalle entre deux fréquences f_1 et f_2 étant : $c = 1200 \log_2(\frac{f_2}{f_1})$[5].

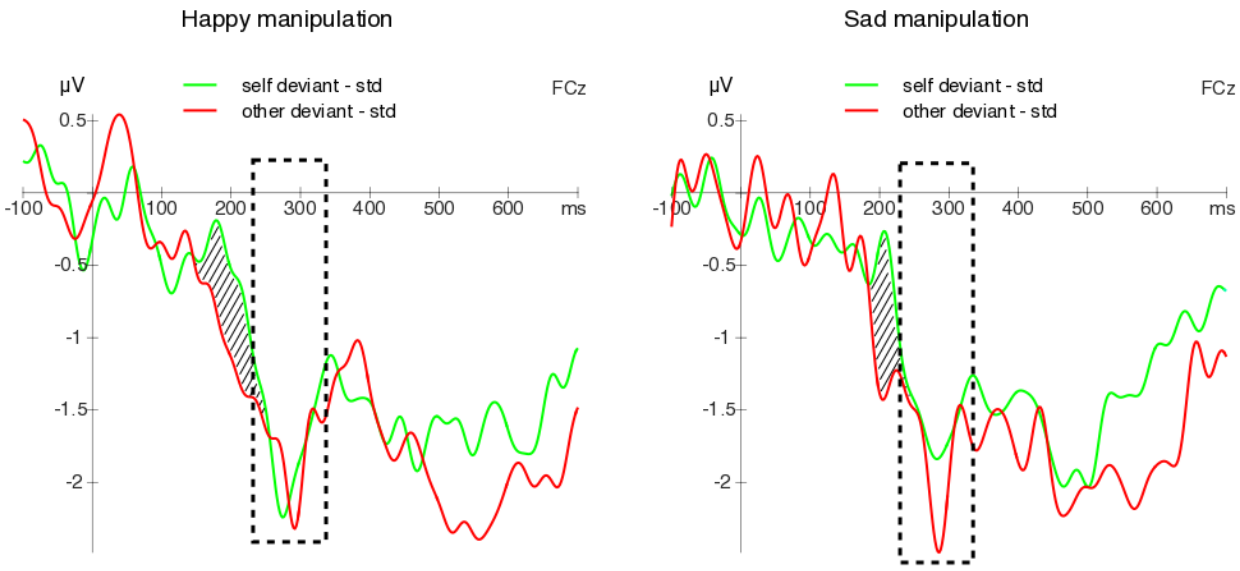

FIGURE 3 – Négativité de déviance (encadrée en pointillé) évoquée par des déviants prosodiques joyeux et tristes, comparées sur des séquences de mots prononcés par le participant ("self") ou par un tiers ("other"). Les déviants prosodiques joyeux et tristes sont traités plus précocement sur la voix d'un tiers (zone hachurée). Figure adaptée de Rachman et al (en préparation).

dans une séquence de stimuli standards. Cette étude a montré une différence d'onset de la réponse de mismatch entre des séquences de mots prononcés par le participant et par un tiers : une variation prosodique sur un mot prononcé par un tiers (manipulé comme joyeux ou triste par le logiciel DAVID) est traité plus précocement que la même variation appliquée sur un mot prononcé par le participant, avec une différence de plus de 30ms (Figure 3). Ce travail, en cours de publication, fait partie du travail de thèse dans l'équipe CREAM de Laura Rachman.

Le présent stage, co-encadré par Jean-Julien Aucouturier et Laura Rachman, s'inscrit dans la continuité de cette série de résultats. D'une part, le fait que les composants MMN indexant le traitement perceptif de la voix du participant et de la voix d'un tiers ne diffèrent pas en fonction de la manipulation émotionnelle joyeuse ou triste suggère que ces processus peuvent être étudiés avec des manipulations acoustiques plus bas-niveau (non spécifiquement émotionnelles) comme la seule modification de la hauteur (qui est commune aux deux manipulations émotionnelles, vers le haut pour joyeux, vers le bas pour triste). D'autre part, le fait que des différences soient visibles au niveau de la négativité de déviance suggère qu'il existe une différence de détection d'erreur différent pour la voix du self par rapport à la voix d'un tiers. (*error monitoring*, voir section 1.3). Nous nous proposons donc, dans ce stage, d'étudier les composants EEG relatifs à la détection d'erreur dans la production de parole suite à un retour auditif modifié par manipulation de pitch (voir section 1.1). L'originalité de ce travail est de se placer dans le cadre encore non étudié de la parole continue (et non de la production de phonèmes isolés, voir section 1.3).

Dans l'état actuel de l'expertise existante dans l'équipe, cette étude implique donc plusieurs contributions :

1. étendre l'outil DAVID pour l'utiliser comme dispositif de manipulation du pitch du retour vocal, avec la possibilité de piloter les paramètres de la manipulation essai par essai et d'indexer les manipulations dans le temps de façon suffisamment précise pour pouvoir l'utiliser dans un paradigme d'EEG ERP.
2. de concevoir une tâche psychophysique de calibration de la difficulté de la tâche de détection d'erreur, en fonction de la taille de la manipulation de pitch

3. d'implémenter un paradigme expérimental EEG applicable dans le cas de la parole continue
4. de collecter les données EEG dans deux premières conditions expérimentales (parole produite vs parole perçue), et de les analyser pour contraster ces conditions.

Au delà du stage, ce dispositif expérimental devra être utilisable pour contraster d'autres conditions, notamment entre parole de soi et parole d'un tiers, et entre signaux parlés et signaux chantés.

Le présent rapport s'organise de la façon suivante : la première partie traitera de l'électroencéphalographie (EEG) et ses techniques, seront présentés par la suite le protocole expérimental, l'acquisition et le traitement des données. Pour finir, une interprétation des résultats obtenus et une discussion sera fournie.

Chapitre 1

Indices EEG du contrôle de la voix

Dans cette partie une brève introduction sur l'électro-encéphalographie et les potentiels évoqués sera fournie. Nous présenterons également le paradigme utilisé et regrouperons l'ensemble des résultats obtenus lors des recherches précédentes en vue de poser les hypothèses et objectifs relatifs à notre étude.

1.1 Paradigme du contrôle de la voix

Lorsque nous produisons une action, nous avons une idée précise de l'effet escompté de cette action. Par exemple, lorsque nous voulons attraper un verre sur une table, nous dirigeons notre main vers un point particulier dans l'espace. Si nous voyons que notre main est trop loin du verre pour l'attraper, nous sommes capables de modifier la trajectoire de celle-ci pour atteindre notre but et attraper ce verre. Cet exemple montre comment notre système nerveux central maintient et ajuste un modèle interne des actions que nous produisons [6]. L'idée d'un contrôle de l'état reçu à chaque action produite peut également s'étendre à un contrôle moteur de la parole [7, 8]. Lors d'une conversation ou lorsque l'on lit un texte par exemple, notre cerveau ainsi que notre appareil sensorimoteur lié au langage (langue, bouche, mâchoire, cordes vocales etc...) nous permet de contrôler notre voix, d'y ajouter des émotions ou encore donner un rythme. Mais que se passe-t-il dans notre cerveau lorsque des aspects des sons que nous produisons habituellement sont modifiés (par exemple, changement de hauteur au cours d'une phrase) ? Pour répondre à une telle question le paradigme principalement utilisé est dit de "retour auditif modifié" (*Altered Auditory Feedback* en anglais) : les participants de l'expérience produisent des sons et entendent en temps réel un retour de ces sons manipulé acoustiquement selon une certaine dimension (voir figure 1.1). Plusieurs études EEG ont été menées autour de cette problématique en utilisant ce paradigme, et plus particulièrement en appliquant un changement de fréquence fondamentale lors de production de phonèmes isolés parlés [9] ou chantés[10]. L'originalité de ce travail de stage est d'étendre ces travaux en mettant en place une expérience EEG impliquant la production de phrases complètes (*running speech*) et un changement de hauteur en temps-réel. Cette étude permettrait d'observer les réactions au niveau cérébral lors de la détection d'un changement de *pitch* dans un contexte réaliste de production de discours.

1.2 Introduction à l'EEG et aux ERP

L'électroencéphalographie (EEG) est la mesure des champs électriques du cerveau à l'aide d'électrodes placées sur la tête. Les champs électriques mesurés sont les résultats de signaux électrochimiques passant d'un neurone à un autre. Lorsque des millions de ces petit signaux traversent simultanément des régions importantes de neurones orientés favorablement, les champs électriques s'additionnent et deviennent assez importants pour être mesurés de l'extérieur du crâne. L'EEG nous permet d'observer

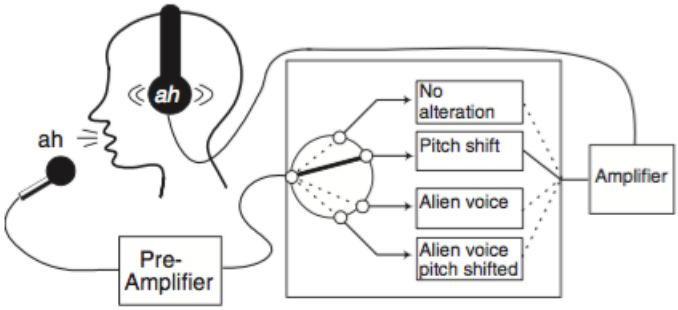

FIGURE 1.1 – Exemple de dispositif pour le paradigme *altered auditory feedback* : le participant produit des phonèmes isolés en s'écoutant au casque. Le retour auditif est manipulé informatiquement pour produire un son différent de celui qui a été produit (ici : la même voix à une hauteur différente, ou une autre voix pré-enregistrée). Figure adaptée de Heinks et collègues [9]

des oscillations neurales définies par des rythmes groupés en un petit nombre de fréquences caractéristiques. Ces rythmes découlent des différentes excitations des populations de neurones et possèdent de complexes modèles qui varient en amplitude, temporalité et fréquence. Le but de l'électrophysiologie cognitive est de comprendre comment et pourquoi ces non-stationnarités sont reliées à des processus cognitifs et des processus de perception[11]. Dans un EEG on peut reconnaître certaines caractéristiques (*features* en anglais) directement liées à des processus cognitifs et sensoriels, on s'intéressera ici en particulier à celles découlant du contrôle de sa propre voix.

Parmi les informations qui peuvent être recueillies par EEG, les réactions ou réponses qui seront analysées sont les potentiels évoqués (PEs, *Event-Related Potentials* ou ERPs en anglais). Ils traduisent la modification du potentiel électrique produit par le système nerveux en réponse à une stimulation (visuelle, auditive...) et ces fluctuations sont directement observables sur les EEGs[12]. Ces potentiels électriques se propagent à travers le cerveau et le crâne quasi-instantanément, ainsi les potentiels enregistrés directement sur le scalp reflètent une activité neurale avec une excellente résolution temporelle. Cependant, ils se propagent également facilement sur la surface du scalp et rendent la localisation précise des ERPs difficile. L'enregistrement de telles données permet d'analyser les composants des ERPs (*components* en anglais), ces composants étant une source de variabilité systématique et fiable dans un ensemble de données ERP[13]. Chaque composant est nommé selon une certaine convention : la première lettre "P" ou "N" indique si le composant est positif ou négatif, s'en suit un nombre indiquant la latence du pic d'amplitude. Par exemple, le composant N400 est négatif et atteint son amplitude maximale 400ms après l'apparition du stimulus. Le nombre figurant dans le composant peut également donner la position du pic dans la forme d'onde. Par exemple, le composant P2 est positif et est le deuxième pic apparaissant dans la forme d'onde (voir figure 1.2). Il arrive que la numérotation des composants soit un peu confuse, par exemple si le composant N1 a son maximum d'amplitude autour de 100ms après le début du stimulus il pourra être noté N1 ou N100, dans tous les cas les notations doivent rester cohérentes et ne pas être mixtes entre les composants.

1.3 Indices relatifs au contrôle de la voix

L'importance de la rétroaction (ou *feedback*) lorsque que l'on parle est due au fait que le rôle premier de la parole soit de produire des sons. Si ces sons sont correctement produits, un tiers pourra les comprendre et les interpréter. Ainsi, le seul moyen de contrôle pour une personne produisant un discours est le son de sa propre voix qui parvient à ses oreilles (autrement dit, son *feedback*). Une théorie a été émise sur les changements sensoriels causés par nos propres actions et est basée sur des modèles

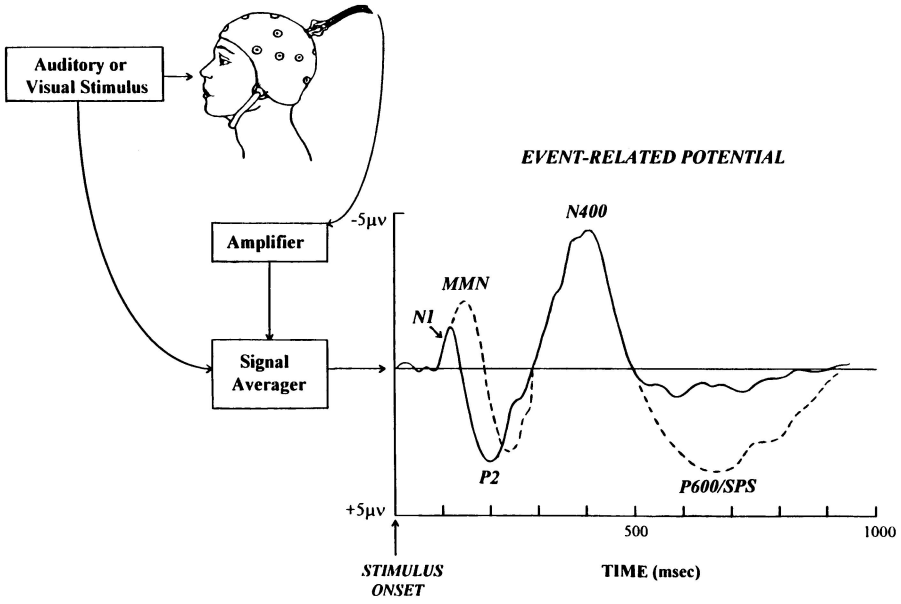

FIGURE 1.2 – Exemple de composants ERP. Le signal EEG enregistré en réponse à un grand nombre de stimulations présentant certaines caractéristiques identiques (par exemple ici, du speech produit ou écouté) est moyenné pour supprimer les variations qui ne sont pas liées au timing du stimulus. L'onde est ensuite analysé en terme d'onde positives (ex. P2, ou 2e pic positif) ou négatives (ex. N1, premier pic négatif, N400, négativité autour de 400ms, etc.) suivant l'*onset* du stimulus. Noter ici l'échelle des amplitudes inversée (composants négatifs vers le haut). Figure adaptée de Lee Osterhout [14]

dits *internal forward*[6, 15]. Selon ces concepts, chaque représentation d'une commande motrice est traduite par un événement sensoriel au niveau cérébral (*efference copy*), qui peut ensuite être comparée au signal perçu ou utilisé pour inhiber la réponse à ce signal (*corrolary discharge*) (voir figure 1.3). Un des phénomènes physiologiques associés au traitement du *feedback* auditif est la "suppression induite par la parole" (*Speaking-induced suppression (SIS)* en anglais) : les réponses corticales auditives liées à la production de discours sont significativement plus faibles que celles qui apparaissant lors de l'écoute passive de ce même discours[8]. Cette "suppression" ne se produit pas lorsque le *feedback* est modifié [8]. Par exemple, lors de la production de voyelle, le cortex supprime l'information sensorielle qui lui est liée : le composant N1 (ou N100 dans ce cas précis) est significativement plus petit lors de la reconnaissance de sa propre voix (lorsque le *feedback* est "correct", non modifié) [9, 16, 17, 18]. L'amplitude de ce composant évolue en fonction de la perturbation (de grandes perturbations du *feedback* entraînent de plus grandes réponses)[19]. Il a été montré que la réduction de ce composant N1 est directement lié au fait que les actions aient été générées par soi (*self-made effect* en anglais)[20], que ce soit dans le domaine auditif[17] ou visuel[21]. On observe d'ailleurs chez des personnes atteintes de maladies telles que la schizophrénie que ces réactions cérébrales sont absentes[22].

Un autre phénomène associé au traitement du *feedback* a pu être montré grâce à des études électrophysiologiques : l'augmentation des réponses due aux perturbations dans le discours (*Speech perturbation response enhancement (SPRE)* en anglais). Les réponses aux perturbations observées sont plus importantes lors de la vocalisation que lors de l'écoute passive[8]. Contrairement au phénomène *SIS*, le *SPRE* est visible non pas dans les premières réactions (par exemple N1), mais dans les réponses tardives : les composants P1 et P2 ont une plus grande amplitude lors de la production[24]. Cette différence peut s'expliquer par le fait que le système nerveux central prédit ce qui va être entendu en se basant sur les mouvements produits pour parler (*efference copy*). Ensuite les prédictions d'erreurs sont traitées par le cortex auditif secondaire pour ajuster et corriger en conséquence la production. Un gain

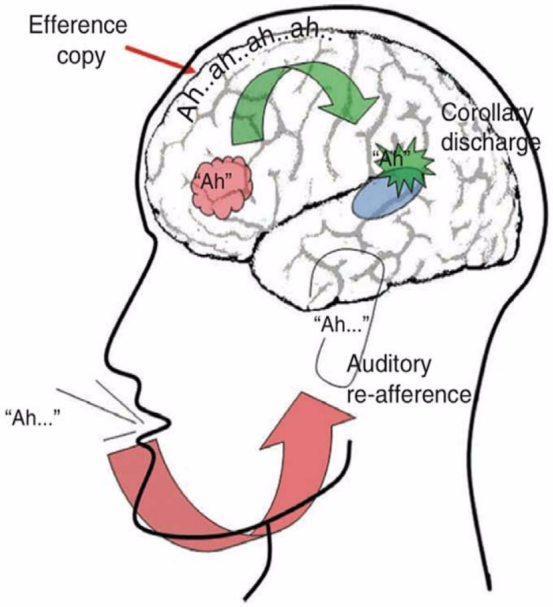

FIGURE 1.3 – Modèle de prédiction interne. La production d'une voyelle [a] s'accompagne d'une copie efférente, qui prédit le signal sonore qui sera perçu en retour. Le signal effectivement perçu (ou *re-afference* auditive) pourra ensuite être comparé à cette copie au niveau du cortex auditif, générant un éventuel signal d'erreur ou une inhibition du traitement (*corollary discharge*). Figure adaptée de Sitek et collègues [23]

est associé à ce processus de correction d'état (le gain de Kalman [6]), ce gain est ajusté pour rendre compte de la corrélation entre les changements perçus dans le *feedback* et les changements du système articulatoire. C'est ici que l'on peut observer une différence entre production et écoute passive. Lors de l'écoute, il n'y a aucun mouvement de l'appareil sensorimoteur et donc le système nerveux central ne peut estimer la corrélation entre la production et ce qui est entendu. L'hypothèse émise est donc que le système nerveux central fixe cette corrélation et donc le gain de correction a un faible niveau, ce qui veut dire que pour une prédiction d'erreur similaire, de plus petites corrections sont générées lors de l'écoute. En d'autres termes, lors de la production, les réponses associées au processus de correction de perturbation sont plus importantes. Celles-ci sont observées plus tardivement car elles sont liées à l'activité du cortex auditif secondaire où les corrections d'état sont générées[8].

Concernant la reconnaissance de sa propre voix, on pourrait ajouter qu'il a récemment été montré que la "détection" de soi découlait de deux processus distincts : *agency assessment* qui représente la causalité des actions et *ownership assessment* qui compare les sons entendus avec ce qui est contenu dans la mémoire[25, 26]. On peut également noter la latéralisation au niveau de l'hémisphère droit pour la reconnaissance de soi, qu'elle soit visuelle[27, 28] ou auditive[29].

1.4 Hypothèses

Nous nous donnons l'objectif de reproduire les résultats précédemment montrés et ainsi étendre les paradigmes communément utilisés sur des phonèmes à des productions plus complexes (*running speech*). Les analyses porteront sur les composants N1, P1 et P2 et peuvent être décomposées en deux parties :

— nous devrions retrouver une petite amplitude pour le composant N1 dans la condition "production" et obtenir un composant N1 d'amplitude significativement plus grande pour la condition

"écoute" (*SIS*), cette différence étant liée au fait que les phrases sont prononcées par les participants eux-même et découle donc un *self-made effect*

— nous examinerons les composants P1 et P2 dans deux conditions différentes (écoute et production) et espérons obtenir des amplitudes plus importantes pour ces composants lors d'un changement de hauteur dans le cas de la production (*SPRE*), traduisant d'un traitement plus haut niveau de l'erreur détectée (*error monitoring*).

De plus, en vue des différentes réactions montrées au niveau cérébral entre production et écoute, nous pourrions également obtenir des différences au niveau comportemental (traduites par une différence de seuil de détection). Cependant, le nombre d'études concernant ce sujet avec de la voix parlée étant peu important, nous ne pouvons émettre de fortes hypothèses quant à la direction de cet effet.

Chapitre 2

Expérimentation

Le contexte de l'étude et les hypothèses étant posés, toute la partie expérimentale peut à présent être développée. Dans cette partie, la mise en place et le déroulement de chaque expérience seront expliqués. Une attention particulière sera portée sur l'acquisition et le traitement des données EEG. Les analyses statistiques qui porteront sur les résultats seront précisées à la fin de ce chapitre.

2.1 Protocole expérimental

N=21 personnes ont été sélectionnées sur la base du volontariat pour participer à cette expérience et ont reçu une compensation à la fin de celle-ci. Trois des 21 participants ont pris part à la phase de pilotage de l'expérience, les analyses statistiques seront conduites sur les résultats des 18 participants restants (10 femmes, 8 hommes). Les participants étaient âgés de 18 à 32 ans, aucun ne présentait de troubles auditifs et quatre d'entre eux étaient musiciens (notons que de futures études pourront comparer les performances de monitoring vocal entre musicien et non-musicien, en utilisant un paradigme identique).

L'expérience se déroule en deux parties distinctes : une partie de production (tâche active) durant laquelle les participants devront lire des phrases et une partie d'écoute passive durant laquelle les participants écouteront les phrases précédemment produites. Dans chaque partie, les participants doivent effectuer deux tâches : une tâche comportementale durant laquelle l'activité cérébrale n'est pas observée et une deuxième tâche au cours de laquelle nous enregistrons les ERPs. Le paradigme d'*altered auditory feedback* sera utilisé tout au long de l'expérience et le changement de hauteur sera dans la direction d'une augmentation du *pitch* (de futures études pourront comparer les résultats obtenus ici avec ceux dans le cas d'une diminution, plutôt qu'une augmentation de hauteur).

2.1.1 Tâche comportementale

Dans cette partie, une procédure psychophysique adaptative a été mise en place pour déterminer le seuil à partir duquel les participants détectent un changement de hauteur dans leur voix. Pour déterminer le seuil de détection des participants, une méthode en escalier (*staircase* ou *up-and-down* en anglais) est utilisée [30]. Les participants munis d'un casque audio et d'un microphone, lisent des paires de phrases et entendent en temps réel leur voix au casque. Pour chaque paire, un changement de hauteur est appliqué aléatoirement sur l'une des phrases et les participants doivent déterminer dans quelle phrase le changement de hauteur a eu lieu (choix forcé à deux alternatives, *2 Alternative Forced Choice* en anglais). Le changement de hauteur initialement appliqué est relativement élevé et celui-ci diminue au cours de l'expérience jusqu'à trouver le seuil à partir duquel le participant détecte un changement de hauteur dans sa propre voix. Pour faire évoluer le changement de hauteur au cours de l'expérience, une procédure *1up-2down* est mise en place [31]. Le changement de hauteur

initialement appliqué est de +200cents, si le participant détecte deux fois de suite le changement de hauteur (i.e. trouve la phrase dans laquelle le changement a été appliqué), la valeur du *pitch-shift* diminue. Inversement, si le participant ne détecte pas le changement de hauteur (i.e. se trompe de phrase), la valeur du *pitch-shift* augmente (voir figure 2.1).

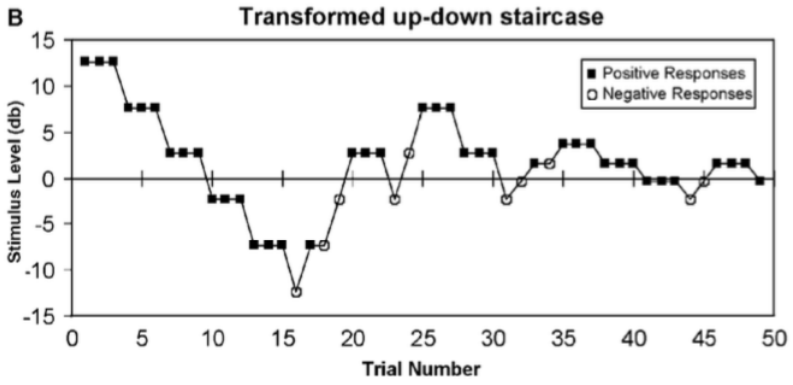

FIGURE 2.1 – Exemple de procédure adaptative de type *staircase* 1up-3down : il faut 3 réponses positives du participant pour baisser le niveau du signal (3-down), rendant ainsi la tâche plus difficile ; il suffit d'une réponse négative pour relever le niveau du signal (1-up). Notre protocole utilise une procédure similaire, avec le critère 1up-2down. Figure adaptée de Leek et collègues [30]

Pour faire converger cette valeur jusqu'au seuil de détection des participants, la valeur du *pitch-shift* évolue selon un certain pas. La valeur du pas diminue pour chaque inversion. Une inversion se produit lorsque la direction du changement hauteur s'inverse (passe d'ascendant à descendant et inversement). Le pas initial utilisé est de 40cents, et celui-ci est divisé par deux à chaque inversion. Le pas minimal est fixé à 2cents. Les valeurs maximales et minimales du *pitch-shift* sont fixées à +200 et +5cents respectivement. Le seuil est déterminé en faisant la moyenne des N dernières valeurs aux inversions. Ici, le nombre d'inversions est fixé à 10 et le seuil est déterminé par la moyenne des valeurs de changement de hauteur aux 5 dernières inversions. Une telle procédure permet de déterminer un seuil pour lequel les participants détectent un changement de hauteur avec une performance de 70.7% [31, 30]. Le changement de hauteur apparaît aléatoirement entre 1 et 1.5s après le début de la production (ou l'écoute) pour diminuer les effets d'anticipation du participant. Le niveau sonore a été ajusté pour chaque participant en fonction de son confort personnel [32]. Pour tous les participants, l'expérience commence avec un entraînement.

Cette procédure sera utilisée dans les deux parties de l'expérience (i.e. production et écoute) afin de déterminer le seuil de détection de chaque participant dans les deux conditions.

2.1.2 Tâche EEG

Après avoir déterminé le seuil de détection de changement de hauteur des participants, la seconde tâche durant laquelle les signaux EEG sont enregistrés est mise en place. Dans cette tâche les stimuli sont présentés un par un aux participants, ils devront lire une phrase dans laquelle un changement de hauteur sera appliqué ou non et devront répondre à la question "Avez-vous perçu un changement de hauteur dans la phrase ?" par "Oui" ou "Non" (tâche de détection). Chaque essai se décompose en quatre parties (voir figure 2.2) :

- une croix de fixation de durée aléatoire (entre 0.7s et 0.9s) pour empêcher l'anticipation du participant
- la phrase est affichée à l'écran pendant une durée fixe de 2s et le participant peut lire la phrase avant la production
- la phrase affichée à l'écran change de couleur et le participant entame la production (durée fixe de 4s)

— le participant fournit la réponse ("Oui" ou "Non") en pressant une touche du clavier
— un écran vide d'une durée fixe de 1.2s qui permet au participant de se reposer.

L'expérience se déroule en 8 blocs de 30 essais, soit 240 essais. Un changement de hauteur est appliqué aléatoirement à la moitié d'entre eux, soit 120 essais avec un changement de hauteur. Les participants liront 60 phrases différentes présentées aléatoirement, chaque phrase sera donc lue deux fois par condition (changement de hauteur ou non).

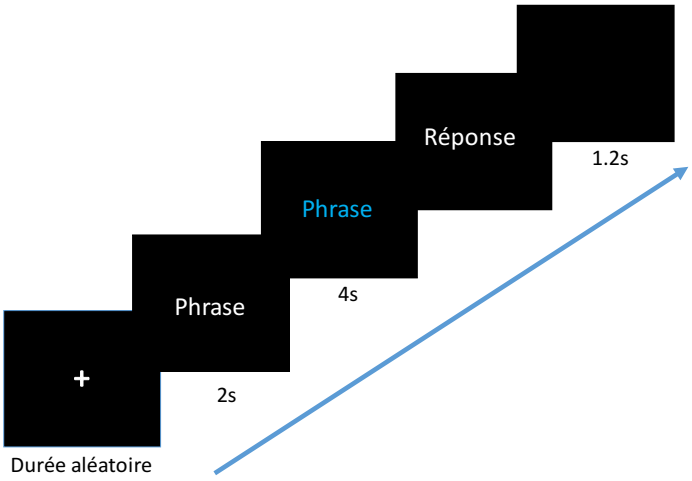

FIGURE 2.2 – Déroulement d'un essai pendant l'enregistrement EEG. Une phrase est affichée à l'écran, puis change de couleur. A ce signal, le participant doit soit lire la phrase à voix haute (condition production), soit écouter l'enregistrement qu'il a précédemment fait de cette phrase (condition écoute). Dans les deux cas, il devra détecter une éventuelle manipulation de hauteur, au seuil, à un instant aléatoire entre 1s et 1.5s pendant l'écoute.

Au cours de cette tâche, le changement de hauteur sera fixé au seuil déterminé précédemment par la procédure adaptative. Ce changement de hauteur fixe nous permet d'avoir un contrôle sur la performance du participant. Comme dans la tâche comportementale, le changement de hauteur apparaît aléatoirement entre 1 et 1.5s après le début de la production (ou l'écoute) pour diminuer les effets d'anticipation du participant. Comme dans la tâche comportementale, l'expérience commence par un entraînement pour chaque participant.

Pour la partie écoute, l'expérience se déroule de la même manière (la phrase étant étendue au lieu d'être lue, à partir du moment ou l'affichage change de couleur à l'écran) et le changement de hauteur est fixé à la valeur de seuil déterminée par la procédure adaptative de la partie écoute.

2.2 Acquisition des données

Pendant toute la durée de l'expérience, les participants étaient installés face à un écran sur lequel les stimuli étaient présentés. L'expérience s'est déroulée depuis un poste informatique Macintosh Mac Mini (Mac OS X 10.11 El Capitan). Les participants étaient munis d'un casque audio Beyerdynamic Dt 770 Pro 250 Ohms et d'un microphone DPA 4088 F Headset tous deux reliés à une carte son RME Fireface UCX. Les changements de hauteur ont été appliqués en temps réel (latence de 19ms) grâce au logiciel DAVID (*Da Amazing Voice Inflection Device*) développé au sein du projet CREAM[3]. L'interface graphique a été réalisée avec Python 2.7 via la librairie Psychopy v1.8[33].

Pour enregistrer l'activité cérébrale des participants et en extraire les ERPs, chacun fut équipé d'un bonnet EASYCAP comportant 64 électrodes actives ('Fp1', 'Fz', 'F3', 'F7', 'FT9', 'FC5', 'FC1', 'C3', 'T7', 'TP9', 'CP5', 'CP1', 'Pz', 'P3', 'P7', 'O1', 'Oz', 'O2', 'P4', 'P8', 'TP10', 'CP6', 'CP2', 'C4',

'T8', 'FT10', 'FC6', 'FC2', 'F4', 'F8', 'Fp2', 'AF7', 'AF3', 'AFz', 'F1', 'F5', 'FT7', 'FC3', 'FCz', 'C1', 'C5', 'TP7', 'CP3', 'P1', 'P5', 'PO7', 'PO3', 'POz', 'PO4', 'PO8', 'P6', 'P2', 'CPz', 'CP4', 'TP8', 'C6', 'C2', 'FC4', 'FT8', 'F6', 'F2', 'AF4', 'AF8', 'Cz') (voir figure 2.4), l'électrode 'Cz' est prise comme référence implicite. De plus nous avons ajouté 4 électrodes passives sur le visage de chaque participant : deux placées à chaque canthus extérieur de l'oeil pour mesurer les mouvements horizontaux des yeux (Horizontal Electro-OcularoGramme/hEOG) et deux autres autres au dessus et au dessous de l'oeil gauche pour mesurer les clignements et mouvements verticaux des yeux (Vertical Electro-OcularoGramme/vEOG) (voir figure 2.3). Pour enregistrer correctement les EEGs/EOGs et obtenir des signaux peu bruités durant l'acquisition, il faut réduire l'impédance de la peau et ainsi créer un pont entre l'électrode et le cuir chevelu ou la peau du visage. Pour réduire cette impédance, il faudra appliquer un gel électrolyte[34] dans chaque électrode à l'aide d'une seringue à bout tronqué (voir figure 2.5). Nous avons visé une impédance maintenue au dessous de 10kΩ pour chaque électrode. Cependant, nous avons accepté un seuil de 20kΩ si la préparation avait duré plus de 45 minutes.

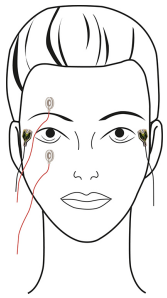

FIGURE 2.3 – Emplacement des électrodes de contrôle des mouvements oculaires (EOG) horizontaux (noires) et verticaux (rouge). Figure adaptée de [35]

FIGURE 2.4 – Bonnet EEG 64 électrodes, en cours de pose sur un participant de l'expérience. La pose et préparation du casque prend au minimum 30 minutes par participant.

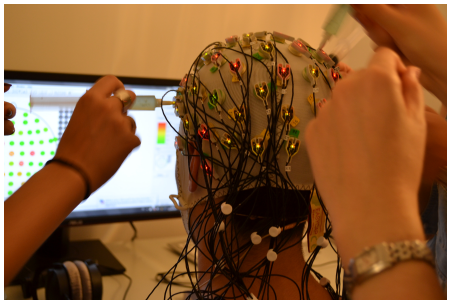

FIGURE 2.5 – Pose du gel électrolyte, électrode par électrode, pour réduire l'impédance de la peau.

Les signaux électriques captés par les électrodes sont de très faibles amplitude (de l'ordre du µV), c'est pourquoi toutes les électrodes sont reliées à un amplificateur afin d'observer correctement les signaux sur l'ordinateur d'acquisition. Ici, un amplificateur actiChAmp (Brain Products) avec une fréquence d'échantillonnage de 500Hz nous permet d'enregistrer des données toutes les 2ms. En plus d'un amplificateur, un détecteur d'événements StimTracker (Cedrus) est utilisé pour détecter et marquer les événements clés de chaque essai pendant l'enregistrement EEG (voir figure 2.6). Il y a trois événements à marquer pour chaque essai :

- le début de l'essai (i.e. l'ouverture du microphone dans DAVID)
- le moment où le participant commence la production (ou l'écoute) (*speech onset* en anglais)
- le moment précis où le changement de hauteur est appliqué.

Au début de chaque essai, le script expérimental écrit en Python envoie un marqueur, via port parallèle, au boîtier StimTracker et celui-ci le transfère à l'ordinateur d'acquisition pour inclusion dans le signal EEG. Le boîtier StimTracker est également branché en série sur la chaîne audio, et génère un autre marqueur dés que le signal sonore dépasse une seuil d'amplitude. Ce marqueur, qui correspond à l'instant où un son sort du casque audio (*speech onset*, lorsque le sujet commence la production - ou l'écoute -), est également transféré à l'ordinateur d'acquisition. Enfin, nous avons étendu (en collaboration avec Marco Liuni, chercheur dans l'équipe PDS) les fonctionnalités du logiciel DAVID pour stocker en temps réel le moment où le *pitch* est modifié - cet instant peut-être ajouté au marqueur de début d'essai pour indexer précisément le moment de la manipulation, qui intervient entre 1 et 1.5s après le *speech onset*.

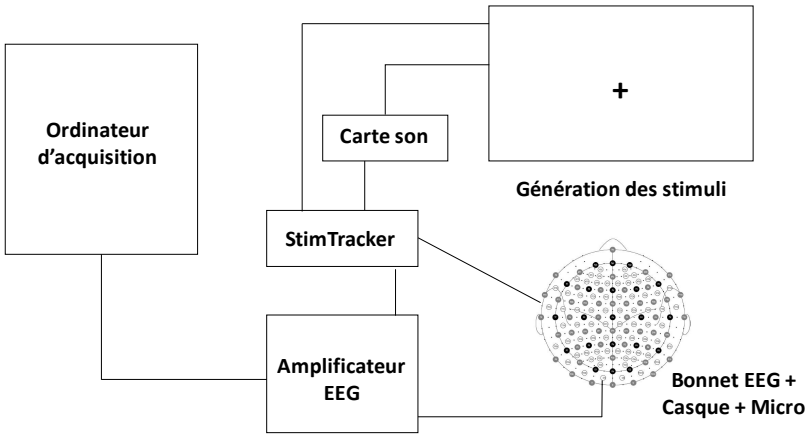

FIGURE 2.6 – Configuration de l'expérience : le dispositif EEG, composé d'un ordinateur d'acquisition, d'un amplificateur de signaux venant du bonnet posé sur le participant, et du boîtier StimTracker, est intégré à la chaîne audio de façon à pouvoir indexer les instants précis du début du *feedback* audio, et de l'*onset* des manipulations de hauteurs appliquées sur celui-ci.

Deux types de données sont enregistrées pour chaque participant : les enregistrements EEG (données temporelles continues) et le fichier comprenant tous les marqueurs envoyés par StimTracker (début de l'essai et *speech onset*) avec leur échantillon correspondant. Les marqueurs indiquant les changements de hauteur délivrés par DAVID sont stockés dans un fichier différent et chaque marqueur correspond à un instant en milliseconde (par opposition à un échantillon, car le logiciel de génération de stimulus DAVID n'opère pas sur la même horloge que l'ordinateur d'acquisition).

2.3 Traitement des signaux récoltés

A la fin de chaque expérience, il a été récolté des données EEG continues de 64 électrodes ainsi que les données continues des EOGs. Avant d'analyser ces données, plusieurs traitements sont appliqués aux signaux. Tout d'abord il faut redéfinir chaque essai présent dans les données, pour cela on utilise les marqueurs enregistrés par l'ordinateur d'acquisition. Chaque essai doit comporter trois marqueurs : le premier marqueur envoyé par python indiquant le début de l'essai, le marqueur du *speech onset* et le marqueur envoyé par DAVID lorsqu'un changement de hauteur a lieu. Il faudra au préalable ajouter les marqueurs de changement de *pitch* au fichier comprenant les autres marqueurs, sans oublier de convertir les instants en échantillons. Sachant que seulement la moitié des essais présentent un changement de hauteur, des marqueurs pour les essais sans changement de hauteur seront crées aux mêmes instants pour chaque phrase correspondante. Ainsi, pour chaque phrase, les ERPs obtenus pour des essais avec et sans changement de hauteur pourront éventuellement être comparés lors des analyses. Une fois tous les marqueurs présents, les enregistrements EEG peuvent être traités. Tous les traitements seront réalisés à partir de la boîte à outils FieldTrip pour Matlab[36].

Une fois les signaux décomposés en essais, un pré-traitement est appliqué : les signaux sont filtrés par un filtre passe-bande de fréquence de coupure 0.5 et 30Hz et sont re-référencés aux électrodes mastoïdes (TP9, TP10) pour réduire les artefacts liés à la parole. Chaque essai peut ensuite être examiné et ceux dont les signaux ne pourront être exploités (amplitudes trop élevées, très bruités) peuvent être manuellement supprimés. Une fois ces essais supprimés, nous pouvons détecter les artefacts visuels (i.e. clignements et mouvements horizontaux, voir figures 2.7 et 2.8 respectivement) à l'aide d'une analyse en composantes indépendantes (*Independent Component Analysis (ICA)* en anglais)[37]. Ici la

méthode fastICA utilisée permet de visualiser les différentes composantes de nos signaux par essai et ainsi les composantes associées aux artefacts visuels peuvent être supprimées.

Pour effectuer les analyses, les données seront décomposées en époques de 600ms, commençant 100ms avant le *speech onset* et jusqu'à 500ms après celui-ci. Un deuxième type d'époques est également considéré : celles-ci autour du changement de hauteur (100ms avant et 500ms après) et pourra être discuté à la fin de l'étude.

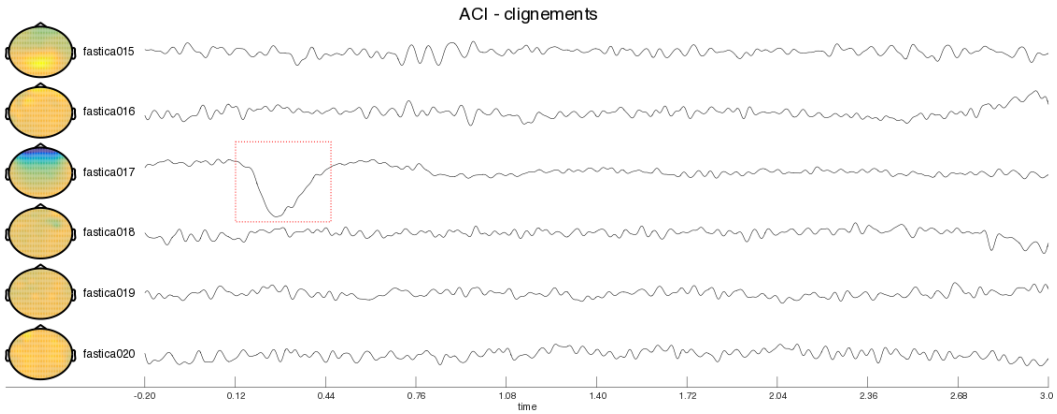

FIGURE 2.7 – Exemples de pré-traitement des données EEG par analyse en composantes indépendantes. Le composant ICA 17 (marqué en rouge) ne présente pas une dynamique temporelle typique de signaux cognitifs, mais est typique des potentiels résultants d'un clignement des yeux (blink). Cette composante doit donc être supprimée, et le signal reconstruit sans la prendre en considération.

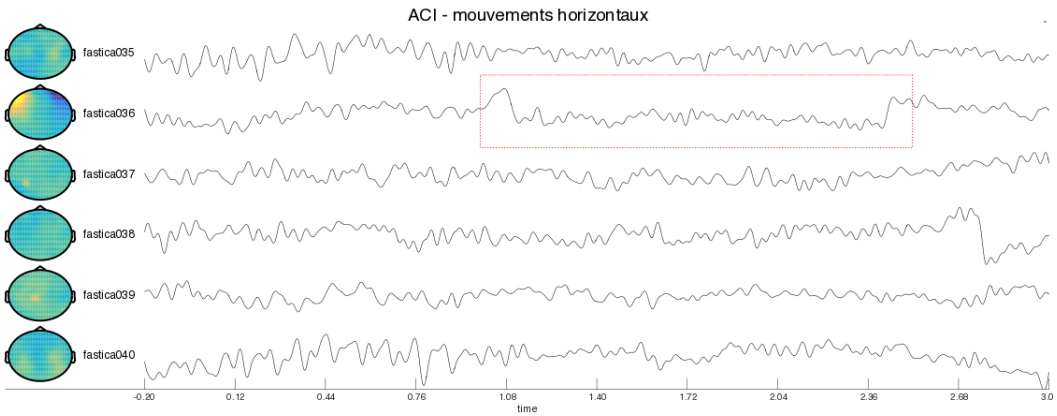

FIGURE 2.8 – Exemples de pré-traitement des données EEG par analyse en composantes indépendantes. Le composant 36 (marqué en rouge) présente un shift tonique typique d'un mouvement horizontal des yeux, et doit donc être supprimé.

2.4 Analyses statistiques utilisées

Les époques ont été moyennées sur 18 participants (*grand average* en anglais) pour les conditions "production" et "écoute". Les amplitudes du composant N1 ont été sélectionnées depuis 15 électrodes fronto-centrales où l'effet N1 est le plus distinct : F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4[23]. Les maxima d'amplitudes pour les composants N1 des participants ont été extraits et moyennés dans une fenêtre de 30ms entre 100 et 130ms après le *speech-onset*. Ces maxima ont ensuite été soumis à une analyse de variance à trois facteurs sur mesures répétées (*three-way repeated measures analysis of variance (RM-ANOVA)* en anglais, voir encadré 1) : Condition (production vs. écoute), Latéralité (5 lignes de la gauche vers la droite, centrées autour de la ligne médiane), Frontalité (Frontal, F vs. Fronto-central, FC vs. Central, C).

Encadré 1 : **Analyse de variance (ANOVA)**

Avant d'entamer une série d'expériences, il faut élaborer un plan expérimental à partir d'une question de recherche. Puis il faut mettre en place un protocole pour tester la question de recherche, ce qui revient à poser une hypothèse nulle (par exemple, "il n'y a pas de différence significative entre deux conditions expérimentales") ; à définir les conditions expérimentales (variables indépendantes *VI*) ; à définir les mesures (variables dépendantes *VD*) et enfin effectuer l'analyse. Pour l'analyse, il est nécessaire d'utiliser des outils mathématiques inférentiels qui permettent de rejeter ou non l'hypothèse nulle pour une population à partir d'un échantillon de celle-ci. L'ANOVA est donc une aide pour argumenter une question de recherche.
Le traitement statistique de données expérimentales a pour objectif de résumer les données (moyenne, écart type ... : statistiques descriptives), et d'estimer la consistance des effets mis en évidence (différence entre les moyennes : statistiques inférentielles). L'ANOVA est un outil d'analyse statistique inférentielle qui permet d'attester la validation d'observations issues de statistiques descriptives utilisées pour dépouiller les résultats d'une expérience. L'ANOVA consiste à tester l'hypothèse nulle (H0).
Pour interpréter l'effet de la variable indépendante, il y a deux variabilités à considérer : la variabilité inter groupes (différences entre les moyennes des groupes dépendants des valeurs de la VI) et la variabilité intra groupe (différences entres les évaluations des sujets d'un même groupe pour une valeur de la VI). La comparaison de la variabilité inter groupes par rapport à la variabilité intra groupe revient à poser la question : "Est-ce que l'hypothèse nulle est validée ou rejetée ?". Le critère de l'effet de la variable indépendante est défini par l'indice :

$$Fcal = \frac{\text{Variabilité entre les groupes}}{\text{Variabilité inter groupe}}$$

La comparaison de la valeur calculée Fcal à la distribution de Fisher pour les paramètres correspondants aux degrés de liberté respectivement pour les variabilité inter et intra groupes permet ou pas de rejeter l'hypothèse nulle. De manière conventionnelle, l'hypothèse nulle est rejetée de manière significative si $p<0.05$ (le résultat est statistiquement significatif au seuil de signification 5%) et de manière très significative si $p<0.01$. Un indice *Fcal* grand indique un effet de la VI, au contraire plus *Fcal* est proche de 1, "moins la VI a d'effet". En ce qui concerne les facteurs, plusieurs situations sont possibles, le plus souvent dans une expérience avec des sujets on considère des "mesures répétées" ; le facteur sujet S est croisé avec les facteurs de l'expérience. Par exemple, lorsque tous les sujets réalisent l'ensemble des conditions expérimentales pour les facteurs A et B, on considère le plan S*A*B ; lorsque les sujets réalisent partiellement toutes les conditions, seulement pour une modalité du facteur A, on considère le plan à mesures partiellement répétées S(A)*B ; avec A le facteur inter-sujet (*between*) et le B le facteur intra-sujet (*within*).

Chapitre 3

Résultats

Les résultats obtenus à la suite des expériences seront présentés dans ce chapitre. Une discussion portant sur les présents résultats sera proposée au prochain et dernier chapitre.

3.1 Données comportementales

A la fin de chaque tâche comportementale, le seuil de détection de changement de hauteur de chaque participant est calculé. Un seuil différent est donc calculé pour chaque condition : "production" (M=+52cents) et "écoute" (M=+46cents). Un test statistique (*paired t-test*) a été effectué pour comparer les moyennes des seuils récoltés (voir figure 3.1). Il n'y a pas de différence significative (p>0.05) entre les conditions "production" et "écoute" mais une tendance se dégage avec un seuil légèrement plus élevé en moyenne pour la condition "production" (cad qu'il serait marginalement plus difficile de percevoir un changement de hauteur sur sa propre production vocale, que a posteriori, à l'écoute de l'enregistrement de celle-ci).

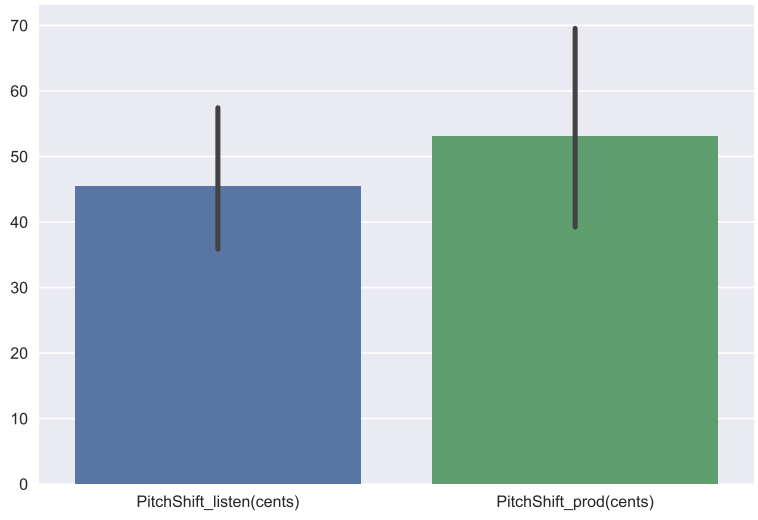

FIGURE 3.1 – Seuils de détection de changement de hauteur (en cents) moyennés sur tous les participants pour les conditions "production" (vert) et "écoute" (bleu)

3.2 Données EEG

3.2.1 Composant N1

Les ERPs ont été enregistrés pendant la production de phrase et l'écoute passive de celles-ci. Les composants N1 ont ensuite été mesurés à partir du *speech-onset* pour les deux conditions et soumis à une analyse de variance. Nous retrouvons bien une suppression significative du composant N1 pendant la production (*Speaking induced suppression*) : $F(1,17)=6.667$, $p=0.019$ (amplitudes moyennées du composant N1 pour la condition production : -1.30µV, pour la condition écoute : -2.74µV), voir figure 3.2. Nous n'obtenons pas d'autres effets ou interactions significatives. Notons que l'on retrouve clairement la localisation fronto-centrale du composant N1 sur la distribution topographique (voir figure 3.4).

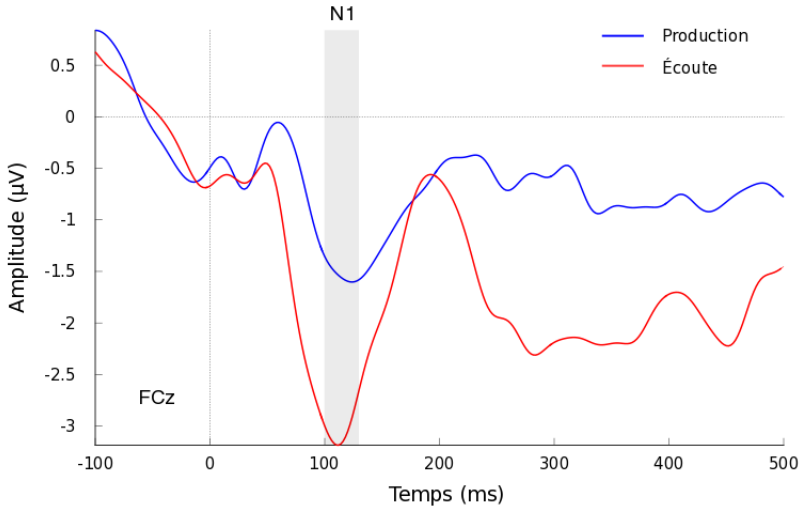

FIGURE 3.2 – *Grand average* ERPs sur 18 participants enregistrés par l'électrode fronto-centrale FCz pour les conditions "production" et "écoute". La partie hachurée représente la fenêtre dans laquelle les maxima du composant N1 ont été extraits. L'origine des abscisses représente le *speech onset*.

3.2.2 Composants P1, P2

Afin d'analyser les composants tardifs, dont P2, les signaux sont ré-alignés sur l'*onset* de la manipulation de pitch appliquée au *feedback*, dans chaque essai. Nous présentons ici le "*grand average*" (sur 17 participants, 1 participant ayant été exclu pour cause d'absence de marqueurs de manipulation). Même si l'on peut constater que l'onde en condition production dévie rapidement de la *baseline*[1], il apparaît qu'il est très difficile d'identifier un composant positif qui soit comparable entre les conditions. Trois composants semblent identifiables dans la condition de production, à 50ms, 100 et 150ms, et 2 seulement dans la condition d'écoute. Le composant à 150ms, qui est plus positif en condition manipulée qu'en condition non-manipulée, et plus grand en condition active que passive, pourrait se comporter comme le SPRE de la littérature. Mais la présence possible d'autres composants dans le voisinage, et l'incertitude sur la localisation temporelle de ceux-ci, ne permet pas de construire une interprétation scientifique robuste sur ces données.

Une explication possible à ce niveau de bruit observé au voisinage de la manipulation de hauteur est dû au paradigme de parole en continu, qui rend la localisation exacte de l'*onset* de la manipulation difficile à mesurer. Même si une solution complète à ce problème n'a pas pu être finalisée au moment

1. Partie du signal précédant le stimulus (avant l'origine de l'axe des abscisses).

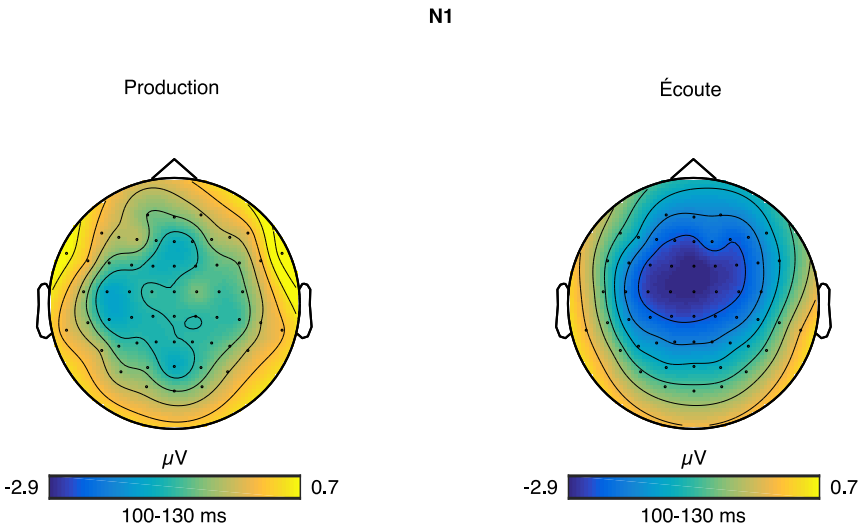

FIGURE 3.3 – Distribution topographique du composant N1 pour les conditions "production" et "écoute". Dans les 2 conditions, le composant est distribué de façon fronto-centrale, en accord avec la littérature.

de la rédaction de ce rapport, nous proposons dans la section suivante une stratégie d'analyse qui pourrait selon nous solutionner ce problème.

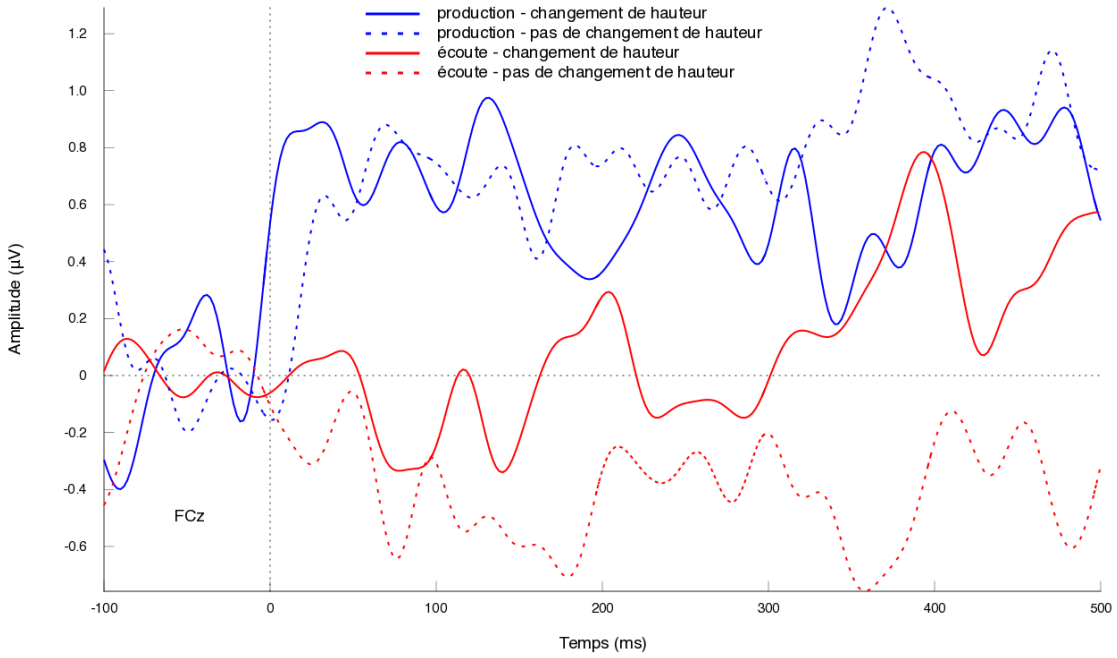

FIGURE 3.4 – *Grand average* ERPs synchronisé sur l'*onset* de la manipulation de hauteur sur 17 participants enregistrés par l'électrode fronto-centrale FCz pour 4 conditions différentes : production avec changement de hauteur, production sans changement de hauteur, écoute avec changement de hauteur, écoute sans changement de hauteur.

Chapitre 4

Discussion

Le présent stage s'est principalement penché sur la question du contrôle de sa propre voix (*self-monitoring* en anglais). Ce contrôle qui se produit de façon automatique lorsque nous parlons est possible grâce au retour auditif que nous entendons à chaque fois que nous produisons des sons. Ce retour est ce qu'il y a de plus important car c'est sur celui-ci que se base notre cerveau pour détecter une erreur dans un discours : en fonction des commandes motrices produites (mouvements de l'appareil sensori-moteur), l'état attendu suite à ces commandes est prédit par le cerveau et comparé avec le *feedback* reçu (*internal forward model*). Plusieurs études ont été mises en place pour étudier les phénomènes liés au *self-monitoring* en utilisant des paradigmes dits *altered auditory feedback*. Un paradigme similaire a été utilisé ici en modifiant la hauteur (*pitch*) de la voix. Les vocalisations mises en jeu dans la plupart des études se réduisant à des phonèmes ou des mots, nous avons voulu mettre en place des nouvelles méthodes pour étendre ces paradigmes à des discours plus complexes comme la production de phrases, plus proches de ce que nous produisons tous les jours pour communiquer (*running speech*). Nous avons également voulu apporter une nouvelle dimension aux paradigmes existant en ajoutant une étape psychoacoustique avant l'étude électro-physiologique. Ce test psychophysique nous a permis de calculer les seuils de détection de changement de hauteur des participants dans leur propre voix et ainsi également contrôler le taux de détection des participants pendant l'EEG. Contrairement à certaines études, le changement de hauteur est appliqué aléatoirement pendant la production (ou l'écoute) pour empêcher l'anticipation des participants.

Nos analyses ont pu faire ressortir les résultats précédemment montrés concernant la suppression des réponses corticales liées à la production de sons (*speaking induced suppression*) : le composant N1 est significativement plus petit lors de la production active[17], suggérant un mécanisme de prédiction interne lorsqu'une action est produite. Les données récoltées à la fin de l'expérience ne nous permettent pas de montrer une quelconque modulation de ce composant en fonction du changement de hauteur[18] car les modifications du *feedback* apparaissait plus tardivement que ce potentiel (aléatoire entre 1s et 1.5s post *onset*) et non au début des vocalisations. Ce résultat n'est donc pas de nature à nous donner des indications concernant le traitement de l'erreur reçu dans le *feedback*, mais nous confirme les idées des modèles dits *internal forward*.

Concernant l'analyse des composants tardifs (P2), le traitement des données est différent de celui appliqué pour extraire un phénomène *SIS*. Le composant N1 a été extrait à partir du *speech onset* (voir partie 2.3) mais pour extraire les composants tardifs liés au traitement d'erreur (*SPRE*), il faut observer les signaux autour du changement de hauteur. Notre première analyse, indexée sur l'*onset* de la manipulation enregistré par le logiciel DAVID, a montré que les réponses neuronales sont trop bruitées autour de ces instants pour pouvoir mesure un potentiel P1 ou P2 de façon précise.

Notre interprétation de ce problème est la suivante : les sons produits par les participants étant des phrases de plusieurs mots, un changement de hauteur peut se produire pendant un moment de silence (ou même pendant une consonne) et donc ne pas être perçu par le participant à l'instant de la manipulation, mais plus tard. Avec un taux syllabique typique de 4 Hz (250ms), un bruit d'*onset*

de 50 ou 100ms n'est donc pas impossible. Dans ces conditions, le "moyennage" de signaux EEG non-alignés rend l'analyse ERP impossible. Pour réduire le bruit possiblement induit par ce phénomène dans nos données, un traitement sur les fichiers audio récoltés pendant la production a été effectué pour détecter les parties voisées du discours et ainsi trouver la partie voisée la plus proche d'un marqueur placé pendant un silence (ou consonne) et l'y déplacer. Pour retrouver les parties voisées, nous avons utilisé un algorithme de superVP (AudioSculpt v3.3.8) qui mesure l'harmonicité de notre signal trame par trame en utilisant l'indice de confiance d'estimation du pitch [38]. Autrement dit, l'harmonicité du signal audio produit par les participants est comparée à l'harmonicité d'un signal parfaitement harmonique (rapport), on obtient donc un indice (ou score) entre 0 et 1, 1 étant un signal parfaitement harmonique. Nous avons fixé un cet indice à un seuil de 0.8, toutes les parties du signal avec une harmonicité au dessus de ce seuil se verront attribuer un score de 1, toutes les parties au dessous se verront attribuer un score de 0 (voir figure 4.1). Jusqu'à présent, nous avons procédé à ce traitement sur 15 participants. En moyenne, les marqueurs de 46.7 essais (écart-type = 9.7) par sujet doivent être déplacés. Le décalage moyen est de 67.5ms (écart-type = 12.6ms).

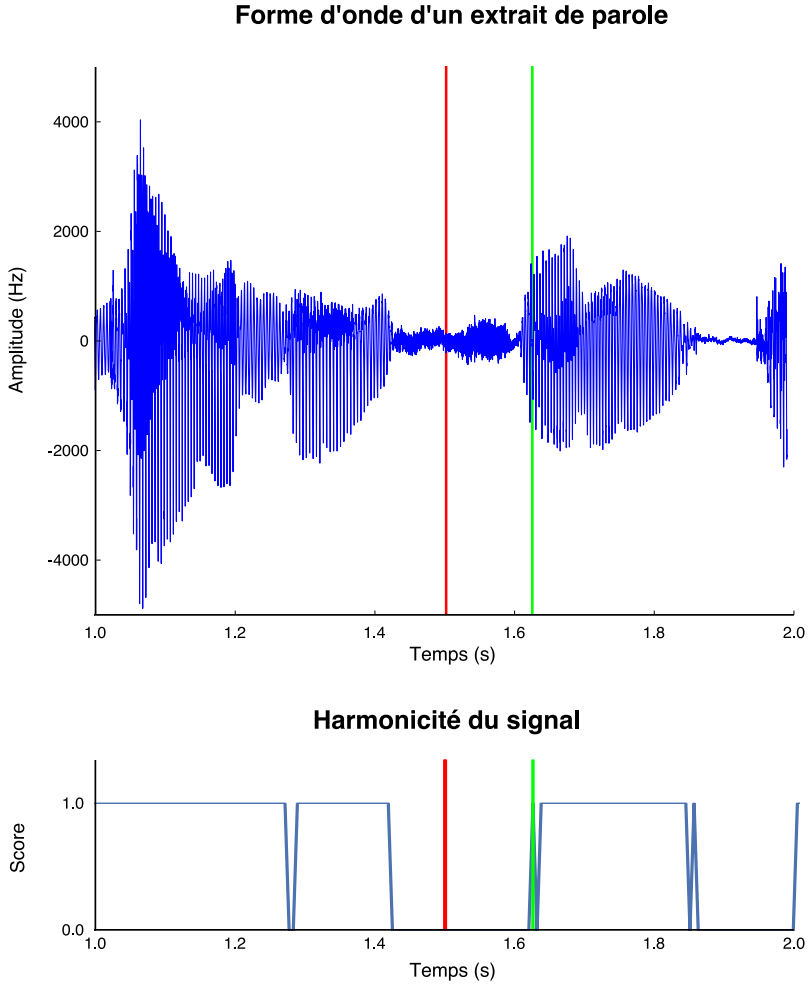

FIGURE 4.1 – Exemple de recalage de marqueur d'événement par analyse acoustique de l'harmonicité du signal : le marqueur original du changement de hauteur (en rouge) placé sur une consonne/silence est déplacé sur la partie voisée la plus proche (en vert).

Une fois tous les marqueurs correctement déplacés sur des parties voisées du signal, nous pourrons réitérer l'analyse des ERPs et espérer en extraire des composants tels que P1,P2 traduisant d'un traitement d'erreur en cas de changement de hauteur dans le *feedback* (*SPRE*). Nous devrions obtenir des composants P2 plutôt élevés, en effet l'amplitude de ce composant semble être modulée en fonction de l'erreur détectée[19] : une erreur trop importante (i.e. le *feedback* reçu ne peut correspondre à l'action

motrice produite) induit de plus petites amplitudes pour le composant P2 comparées à celles que l'on peut observer lorsque le *feedback* est modifié plus subtilement. Il est donc possible pour nous d'obtenir des amplitudes relativement grandes pour le composant P2, le changement de hauteur étant fixé au seuil de détection.

Une fois ce problème de la localisation précise des événements solutionné, possiblement par l'analyse acoustique proposée ci-dessus, le paradigme que nous avons mis en place pourra permettre plusieurs analyses complémentaires. D'une part, le fait que les participants participent à une tâche active, de détection de manipulation essai-par-essai, permettra de comparer les réactions neuronales en fonction des réponses des participants. Nous pourrions comparer les signaux obtenus pour les détections correctes (*Hits* en anglais) avec les détections manquées (*Misses* en anglais) et ainsi observer la présence ou non de certaines réactions en fonction de la réponse fournie. D'autre part, nous pourrons également examiner les fichiers audio enregistrés (tels qu'ils ont été produits par les participants, et non tels qu'ils ont été entendu) et vérifier si une compensation suite au changement de hauteur est visible lors de la production[39, 40]. En cas de compensation, les signaux EEG pourront être analysés pour tenter d'en extraire les mécanismes neuraux qui lui sont liés. Enfin, les résultats obtenus pour la tâche comportementale pourront être plus profondément analysés. Dans l'état actuel des résultats, les différences inter-participants semblent assez élevées et ne nous permettent pas de tirer de conclusion quant à une différence précise entre les conditions production et écoute. Entre autres, une analyse sur les capacités perceptives des sujets (indice *d'* utilisé en théorie de détection du signal[41]) pourra être menée pour tenter d'expliquer les différentes réactions neuronales observées entre production et écoute.

Au delà des données déjà collectées ici, notre paradigme ici centré sur sa propre voix pourra s'étendre à la voix de l'autre. Nous pourrions ainsi comparer les mécanismes et traitements neuraux de la voix d'un tiers avec sa propre voix dans le cas d'un discours continu. De même nous pourrions étendre notre étude à la voix chantée et compléter les résultats d'études précédemment menées sur les traitements de la justesse de la hauteur sonore lors du chant [42, 43] et les comparer avec ceux de la voix parlée. Un étude suggère notamment que la pratique musicale requiert plus de précision dans le traitement auditif[44] et cette hypothèse pourrait servir de base à une future étude confrontant voix parlée et voix chantée, comparant par exemple deux groupes de musiciens et de non-musiciens ou corrélant les résultats des participants à un indice d'habilité musicale comme le PROMS [45] ou le Goldsmith musical sophistication index [46].

Conclusion

L'étude présentée dans ce rapport a donc porté sur le contrôle de sa propre voix et plus précisément lors d'un discours continu. Nous avons voulu étendre les paradigmes existant, traitant du même sujet sur la production de phonèmes, à la production de phrases. Pour ce faire, l'outil DAVID a été adapté aux paradigmes ERP : ce logiciel d'habitude utilisé pour modifier le contenu émotionnel en temps réel a été utilisé ici pour provoquer des changements de hauteur et de nouvelles fonctions ont été ajoutées pour récolter les instants précis de ces changements de hauteur en temps réel (ces instants étant indispensables à l'analyse des données EEG). Deux expériences ont ensuite été conçues : une expérience comportementale qui nous a permis de récolter les seuils de détection de changement de hauteur des participants dans deux conditions ("production" et "écoute") et une expérience électrophysiologiques qui nous a permis de récolter les signaux EEG des participants dans deux conditions ("production" et "écoute") pour en extraire les ERPs.

Les résultats comportementaux n'ont pas montré de différence significative entre les seuils de détection de changement de hauteur pour les conditions "écoute" et "production" (comparables, à +50cents). Des analyses complémentaires des résultats sont à prévoir pour expliquer les différences observées entre ces deux conditions. L'analyse des ERPs a révélé la présence d'une suppression significative du composant N1 pour la condition "production" (*SIS*). Cette suppression traduit la prédiction interne suite à une action produite par soi et confirme les hypothèses des modèles dits *internal forward*. D'autre part, nous avons montré que le paradigme de parole continue rend difficile l'analyse traditionnelle des composants tardifs, probablement à cause de la difficulté de localisation des événements, non synchronisée avec le taux syllabique des participants. Même si une solution complète à ce problème n'a pas pu être finalisée au moment de la rédaction de ce rapport, nous proposons une procédure de recalage des événements grâce à une analyse d'harmonicité du signal sonore qui pourrait selon nous résoudre cette difficulté. Une fois cette difficulté surmontée, nous espérons répliquer les résultats montrés précédemment, à savoir une amplitude significativement plus grande des composants P2 (et/ou P1) lors d'un changement de hauteur pendant la production[24].

L'étude menée au cours de ce stage offre plusieurs perspectives. D'abord, l'amélioration du logiciel DAVID, et son adaptation au paradigme ERP de retour auditif modifié pourra faire l'objet d'une publication dans un article de méthode (dans un revue comme *Neuroscience Methods* ou *Behaviour Research Methods*). En effet, il n'existe pas encore d'outil code source ouvert permettant la modification de hauteur de *feedback* en temps-réel. De plus, lorsque les analyses des données EEG seront terminées la publication d'un article sera également envisagée, nos résultats ayant prouvé la faisabilité d'études EEG dans un contexte plus réaliste de discours continu. Enfin, le paradigme expérimental actuellement menée sur sa propre voix pourra s'étendre à l'étude du traitement de la voix d'un tiers (ou moins dans le cas de la simple écoute) afin d'évaluer et comparer les différences de traitement entre sa propre voix et celle des autres. Une dimension musicale pourra également être explorée en confrontant voix chantée et voix parlée. Dans cette perspective, le travail mené au cours de ce stage se poursuivra avec un contrat d'ingénieure d'études au sein du projet CREAM jusqu'à la fin de l'année.

Bibliographie

[1] www.ircam.fr.

[2] J.J. Aucouturier, P. Johansson, L. Hall, R. Segnini, L. Mercadié, and K. Watanabe. Covert digital manipulation of vocal emotion alter speakers' emotional state in a congruent direction. *Proceedings of the National Academy of Science*, 113(4), 2016.

[3] L. Rachman, M. Liuni, P. Arias, A. Lind, P. Johansson, L. Hall, D. Richardson, K. Watanabe, S. Dubal, and J-J. Aucouturier. DAVID : An open-source platform for real-time transformation of infra-segmental emotional cues in running speech. *Behavior Research Methods*, pages Advance online publication, https ://doi.org/10.3758/s13428–017–0873–y, 2017.

[4] J-D. Bagot. *Information, sensation et Perception.* Armand Colin, 1999.

[5] R. Behroozmand, O. Korzyukov, and C.R. Larson. Effects of voice harmonic complexity on erp responses to pitch-shifted auditory feedback. *Clinical Neurophysiology*, 122 :2408–2417, 201.

[6] D.M. Wolpert, Z. Ghahramani, and M.I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232) :1880–1882, 1995.

[7] John F Houde and Srikantan S Nagarajan. Speech production as state feedback control. *Frontiers in human neuroscience*, 5, 2011.

[8] J.F. Houde and E.F. Chang. The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, 33 :174–181, 2015.

[9] T.H. Heinks-Maldonado, D.H. Mathalon, M. Gray, and J.M. Ford. Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42 :180–190, 2005.

[10] S. Patel, A. Lodhavia, S. Frankford, S. Korzyukov, and C.R Larson. Vocal and neural responses to unexpected changes in voice pitch auditory feedback during register transitions. *Journal of Voice*, 30(6) :772.e33–772.e40, 2016.

[11] M.X. Cohen. Where does eeg comes from and what does it mean ? *Trends in Neurosciences*, 40(4) :208–218, 2017.

[12] E.S. Kappenman and S.J. Luck. Best practices for event-related potential research in clinical populations. *Biological Psychiatry : Cognitive Neuroscience and Neuroimaging*, 1 :110–115, 2016.

[13] S.J. Luck. *An Introduction to the Event-Related Potential Technique - Second Edition.* The MIT Press, 2014.

[14] http://faculty.washington.edu/losterho/erp_tutorial.htm.

[15] R.C. Miall and D.M. Wolpert. Forward models for physiological motor control. *Neural Networks*, 9(8) :1265–1279, 1996.

[16] J. Graux, M. Gomot, S. Roux, F. Bonnet-Brilhault, V. Camus, and N. Bruneau. My voice or yours ? an electrophysiological study. *Brain Topography*, 26 :72–82, 2013.

[17] P. Baess, J. Horváth, T. Jacobsen, and E. Schröger. Selective suppression of self-initiated sounds in an auditory stream : An erp study. *Psychophysiology*, 48 :1276–1283, 2011.

[18] R. Behroozmand and C.R. Larson. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, 12 :54–64, 2011.

[19] N.E. Scheerer, J. Behich, H. Liu, and J.A. Jones. ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience*, 240 :176–185, 2013.

[20] S.K. Poonian, J. McFadyen, J. Ogden, and R. Cunnington. Implicit agency in observed action : Evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of Cognitive Neuroscience*, 27(4) :752–764, 2015.

[21] A. Gentsch, S. Schütz-Bosbach, T. Endrass, and T. Kathmann. Dysfunctional forward model mechanisms and aberrant sense of agency in obsessive-compulsive disorder. *Biological Psychiatry*, 71(7) :652–659, 2012.

[22] J.M. Ford, D.H. Mathalon, S. Kalba, S. Whitfield, W.O. Faustman, and W.T. Roth. Cortical responsiveness during talking and listening in schizophrenia : An event-related brain potential study. *Society of Biological Psychiatry*, 50 :540–549, 2001.

[23] K.R. Sitek, B.H. Mathalon, B.J. Roach, J.F. Houde, C.A. Niziolek, and J.M. Ford. Auditory cortex processes variation in our own speech. *Plos One*, 8(12) :e82925, 2013.

[24] R. Behroozmand, L. Karvelis, H. Liu, and C.R. Larson. Vocalization-induced enhancement of the auditory cortex responsiveness during f0 feedback perturbation. *Clinical Neurophysiology*, 120 :1303–1312, 2009.

[25] D. Hubl, R.C. Schneider, M. Kottlow, J. Kindler, W. Strik, T. Dierks, and T. Koenig. Agency and ownership are independent components of 'sensing the self' in the auditory-verbal domain. *Brain Topography*, 27 :672–682, 2014.

[26] O. Korzyukov, A. Bronder, Y. Lee, S. Patel, and C.R. Larson. Bioelectrical brain effects of one's own voice identification in pitch of voice auditory feedback. *Neuropsychologia*, 101 :106–114, 2017.

[27] J.P. Keenan, M.A. Wheeler, G.J.Jr Gallup, and A. Pascual-Leone. Self-recognition and the right prefrontal cortex. *Trends in Cognitive Sciences*, 4(9) :338–344, 2000.

[28] J.P. Keenan, A. Nelson, M. O'Connor, and A. Pascual-Leone. Self-recognition and the right hemisphere. *Nature*, 409 :305, 2001.

[29] C. Rosa, M. Lassonde, C. Pinard, J.P. Keenan, and P. Belin. Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition*, 68 :204–214, 2008.

[30] M.R. Leek. Adaptive procedures in psychophysics. *Perception & Psychophysics*, 63(8) :1279–1292, 2001.

[31] H. Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2.2) :467–477, 1971.

[32] E. Globerson, N. Amir, O. Golan, L. Kishon-Rabin, and M. Lavidor. Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception & Psychophysics*, 75 :1799–1810, 2013.

[33] Jonathan W Peirce. Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1) :8–13, 2007.

[34] Minimizing impedances and using the right electrolyte gel saves time and makes research more effective, *Brain Products Release ; October 2013, Volume 48*.

[35] `http://pressrelease.brainproducts.com/eog_actichamp/`.

[36] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. Fieldtrip : open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011 :1, 2011.

[37] A. Delorme, T. Sejnowski, and S. Makeig. Enhanced detection of artifacts in eeg data using high-order statistics and independant component analysis. *NeuroImage*, 34 :1443–1449, 2007.

[38] M. Liuni and A. Roebel. Phase vocoder and beyond. *Musica, Tecnologia*, 7 :73–120, 2013.

[39] T. A. Burnett, M. B. Freedland, C. R. Larson, and T. C. Hain. Voice f0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6) :3153–3161, 1998.

[40] R. Behroozmand, O. Korzyukov, L. Sattler, and C.R. Larson. Opposing and following vocal responses to pitch-shifted auditory feedback : Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America*, 132(4) :2468–2477, 2012.

[41] N.A. Macmillan and C.D. Creelman. *Detection theory : A user's guide - Second edition.* Lawrence Erlbaum Associates, Inc., 2005.

[42] P. Larrouy-Maestri, Y. Lévêque, D. Schön, A. Giovanni, and D. Morsomme. The evaluation of singing voice accuracy : A comparison between subjective and objective methods. *Journal of Voice*, 27(2) :259.e1–259.e5, 2013.

[43] J. Merrill and P. Larrouy-Maestri. Vocal features of song and speech : Insights from Schoenberg's Pierrot Lunaire. *frontiers in Psychology*, 8 :Article 1108, 2017.

[44] A.D. Patel. The opera hypothesis : assumptions and clarifications. *Annals of the New York Academy of Sciences*, 1252 :124–128, 2012.

[45] Lily NC Law and Marcel Zentner. Assessing musical abilities objectively : Construction and validation of the profile of music perception skills. *PloS one*, 7(12) :e52508, 2012.

[46] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. The musicality of non-musicians : an index for assessing musical sophistication in the general population. *PloS one*, 9(2) :e89642, 2014.