La mie de pain n'est pas une amie : une étude EEG sur la perception de différences infra-phonémiques en situation de variations.

Stéphane Pota[1,2] *Elsa Spinelli*[1] *Véronique Boulenger*[3] *Emmanuel Ferragne*[3]
Léo Varnet[2] *Michel Hoen*[2] *Fanny Meunier*[2]
(1) Laboratoire de Psychologie et NeuroCognition, CNRS UMR5105, Grenoble
(2) Centre de Recherche en Neurosciences de Lyon, INSERM U1028, CNRS UMR5292, Bron
(3) Laboratoire Dynamique du Langage, CNRS UMR5596, Lyon
stephane.pota@gmail.com

RESUME

Nous avons examiné les corrélats électrophysiologiques de la sensibilité des auditeurs aux indices acoustiques fins en condition de variabilité intra-locuteur dans le but de tester la pertinence de tels indices durant le traitement de la parole. Pour ce faire, une version modifiée du paradigme Oddball a été utilisée avec pour stimuli des syllabes homophones telles que *la* et *l'a* et dans une seconde expérience des séquences plus longues telles que *la mie* et *l'amie*. Le principal résultat de cette étude a été l'observation d'une négativité de discordance (MMN) pour les déviants homophones. Le système de perception de la parole est par conséquent sensible aux différences infra-phonémiques entre des séquences homophones malgré le contexte de variabilité de la parole. Les indices acoustiques fins sont donc assez robustes pour pouvoir jouer un rôle dans le traitement de la parole.

ABSTRACT

Robustness of fine acoustic cues and Speech variability: a Mismatch Negativity study

We examined electrophysiological correlates of listener's sensitivity to fine acoustic cues in intra-speaker variability conditions in order to test the relevance of such cues for the speech perception system. For this purpose, a modified oddball paradigm has been used with syllables such as French homophones *la* and *l'a*, and in a second experiment with longer sequences such as *la mie* and *l'amie*, both /lami/. The main result of this study was the observation of a mismatch negativity (MMN) for homophone deviants. Speech perception system is thus sensitive to subphonemic differences between homophone sequences despite the speech variability. Fine acoustic cues are robust enough to play a role in speech processing.

MOTS-CLES : Indices acoustiques fins, Mismatch Negativity, traitement de la parole
KEYWORDS: Fine acoustic cues, Mismatch Negativity, speech processing

1 Introduction

Afin de comprendre la parole, les auditeurs doivent faire le lien entre l'information sensorielle provenant de l'input acoustique et les entrées lexicales stockées en mémoire à long terme. Deux problèmes majeurs sont rencontrés lors de la reconnaissance de la parole : la continuité et la variabilité de son signal acoustique. Parce que ce signal est

continu, les auditeurs doivent segmenter le flux afin d'identifier les mots. Par exemple, le phénomène d'élision rend certaines séquences phonologiquement ambiguës (comme *l'amie* vs *la mie*) : une segmentation correcte est nécessaire pour une bonne compréhension. Bien que ces séquences soient homophones (i.e., dont la transcription phonémique est identique, ici /lami/), il existe tout de même de légères différences acoustiques entre les membres de telles paires comme la montée initiale de fréquence fondamentale (*F0*), caractéristique des débuts de mots de contenus (elle apparaît ici au début de *l'amie* et de *mie*). Spinelli et al. (2010) ont montré au niveau comportemental que, pour une production donnée, ces indices acoustiques fins pouvaient être pertinents lors de la segmentation des mots par les auditeurs.

Bien qu'il semble maintenant admis que certains indices acoustiques soient utilisés en temps réel pour influencer la reconnaissance des mots par les auditeurs, certaines questions importantes restent pour l'instant sans réponse. Il reste notamment à établir si ces indices sont suffisamment robustes pour être utilisés dans un contexte de productions multiples. En effet, il s'avère que les productions d'un même mot, d'une même séquence, par un même locuteur diffèrent les unes des autres (variabilité intra-locuteur). Cependant les auditeurs semblent garder une trace des probabilités de distribution des signaux acoustiques qui leur sont associées : ils ne sont donc pas seulement sensibles aux informations acoustico-phonétiques, mais aussi à leurs probabilités de distribution. Par exemple, il a été montré que des informations acoustico-phonétiques telles que le VOT (Voice Onset Time) sont capables de restreindre l'activation d'une seule des deux langues pour un auditeur bilingue (Ju et Luce, 2004).

Si certaines caractéristiques acoustiques constituent des indices robustes, notamment pour la segmentation du flux continu, alors elles devraient être disponibles d'une production à l'autre malgré la variabilité du signal et le système neural devrait être sensible à ces différences. La présente étude a donc pour but de tester la robustesse de certains indices acoustiques fins, différenciant des séquences homophones comme *l'amie* vs *la mie*, en conditions de productions multiples.

Afin d'éviter toute focalisation particulière de l'attention des auditeurs sur la forme des mots (ce qui est généralement le cas dans les études comportementales), nous avons ici utilisé une tâche passive couplée à l'enregistrement de potentiels évoqués (PEs). Nous nous sommes intéressés plus particulièrement à une composante majeure des PEs auditifs, associée à la détection de tout changement inattendu dans certains aspects réguliers d'un flux auditif continu: la négativité de discordance (en anglais **MMN** ou Mismatch Negativity ; Näätänen & Alho, 1995). La MMN est une négativité fronto-centrale avec un pic entre 150 et 250 ms après l'apparition du stimulus. Ce PE est obtenu classiquement dans un paradigme dit 'Oddball' au cours duquel un son rare (le *déviant*) apparaît dans une série de stimuli plus fréquents (les *standards*), et cela, indépendamment de l'attention du sujet ou de la tâche proposée. Des études ont montré l'apparition d'une MMN pour des phonèmes déviants, alors même que les phonèmes standards étaient issus de multiples productions (par différents locuteurs, Shestakova et al., 2002), ce qui suggère que ce sont les régularités partagées par les différents standards qui importent. Dans notre étude nous nous sommes intéressés à la perception de différences acoustiques fines infra-phonémiques, pour des séquences homophones. Nous avons examiné les corrélats électrophysiologiques du traitement d'indices

acoustiques fins avec une version modifiée du paradigme Oddball (Brunellière et al., 2010) dans lequel chaque stimulus provenait de productions différentes d'un même locuteur. Dans la première expérience (ExpCV) nous nous sommes intéressés aux syllabes homophones [la#] vs [l#a], et dans la deuxième (ExpMot) à des séquences nominales homophones telles que *la mie* vs *l'amie*.

2 Matériel et méthodes

2.1 Participants, Stimuli et Procédure

Trente-six volontaires de langue maternelle française et âgés de 18 à 24 ans ont participé à notre étude (18 sujets pour chaque expérience : ExpCV : M = 22 ans ; SD = 3 ; 10 femmes; ExpMot : M = 21 ans ; SD = 3 ; 9 femmes). Ils étaient tous droitiers, normoentendants, sans troubles du langage, et sans antécédents de maladies neurologiques.

Les séquences nominales françaises: *la locution*, *l'allocution*, et *l'illocution* ont été extraites de phrases enregistrées par une même locutrice de langue maternelle française (durées moyennes respectives des syntagmes nominaux = 889 ms, 823 ms et 827 ms; normalisation à 65 dB-A). Chacun de ces stimuli provenait de 5 productions différentes de chaque mot. De ces stimuli ont ensuite été extraites les syllabes /la/ ([la#] de *la locution* ou [l#a] de *l'allocution*) et /li/ (durée moyennes respectives = 140 ms, 202 ms et 197ms, *F0* moyennes respectives à la mi-voyelle = 163Hz, 183Hz et 173Hz).

Les passations se sont déroulées dans une salle à isolation électroacoustique. Les sujets, confortablement installés devant un écran d'ordinateur, devaient regarder un film de leur choix sans le son, tout en ignorant les stimuli présentés en stéréophonie via un casque (niveau d'écoute confortable de 65 dB-A). Les sons étaient présentés dans une version modifiée du paradigme *Oddball* (cf. Fig.1), dans lequel une série de quatre standards (identiques mais provenant de productions différentes) était toujours suivie par un stimulus en position test qui pouvait être identique aux standards (**condition identique**, i.e. autre production d'un standard) ou bien être différent, (i) soit en étant un homophone /la/ (**condition homophone** ; parmi 5 productions différentes), (ii) soit un non homophone /li/ (**condition non homophone** ; également parmi 5 productions différentes). Les positions des stimuli ont été pseudo-aléatoirisées et les stimuli étaient séparés par un ISI (Inter-Stimulus Interval) de 500 ms. Chaque expérience a été divisée en deux blocs consécutifs de 1800 stimuli chacun (80% de standards et 20% de stimuli en position test). Les Blocs[l#a] avait pour standard [l#a] et les Blocs[la#], [la#]. L'ordre de présentation des blocs a été contrebalancé entre les sujets.

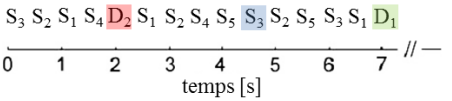

FIGURE 1 – Exemple de séquence.
S_n correspond aux Standards et D_n aux déviants. Encadré en rouge, la *condition homophone*, en vert la *condition non homophone*, et en bleu, la *condition identique*.

2.2 Acquisition et Analyses de l'EEG

L'acquisition EEG a été réalisée avec le système Biosemi à 32 électrodes actives (Electro-Cap International, INC, Ohio, USA ; Biosemi, ActiveTwo, version 5.36) posées sur le scalp des participants selon le système international 10-20. Le signal EEG a été recueilli à une fréquence d'échantillonnage de 2 kHz sur une bande passante de [0.1-400 Hz]. L'enregistrement a été référencé sur la référence commune (CMS) et 1 terre (DRL) directement intégrées au bonnet. L'apparition de chaque son était associée à un marqueur généré par le logiciel Presentation (Neurobs).

Les différentes analyses ont été réalisées avec Fieldtrip (Oostenveld et al., 2011 ; Donders Institute for Brain, Cognition and Behaviour, Pays-Bas). Les données brutes ont d'abord été analysées individuellement. Les conditions de rejet des enregistrements ont été les suivantes : nombre d'électrodes bruitées ≥ 6 ou nombre d'essais bruités par condition $>$ 10%. Ainsi, pour l'ExpCV comme pour l'ExpMot, 16 participants ont fourni des enregistrements de qualité satisfaisante pouvant être inclus dans les analyses ultérieures. Pour les enregistrements conservés, a été réalisée une segmentation automatique calée sur la présentation des stimuli sur une fenêtre temporelle de -200 ms à $+600$ ms pour les syllabes, et de -200 ms à $+900$ ms pour les mots. Une normalisation a été appliquée aux segments en définissant la période pré-stimulus de 200 ms comme ligne de base. Pour l'analyse qualitative des segments, une analyse en composantes indépendantes (ACI) nous a permis d'identifier les artéfacts oculaires et cardiaques, et d'en exclure les composantes avant de recomposer un signal EEG débruité. Les données nettoyées de tout artéfact ont alors pu être moyennées. La visualisation des PEs s'est faite en réalisant au préalable un re-référencement sur la base de l'activité moyenne de deux électrodes externes placées sur les mastoïdes (référence « linked-mastoïd »). Un filtre passe-bas de 20 Hz ainsi qu'un filtre coupe-bande à 50 Hz ont été appliqués offline. Pour chacune des conditions, la procédure classique d'observation des MMNs a été appliquée (par la soustraction PE-déviant moins PE-standard). En accord avec des études précédentes, la réponse MMN est maximale pour la plupart des déviants sur une électrode se rapprochant de Fz dans le système 10-20, cette électrode a donc été choisie pour l'analyse statistique.

Pour chaque bloc et chaque participant, des analyses basées sur un rolling t-test ont été réalisées sur toute la durée des segments afin d'effectuer la comparaison des amplitudes entre les PEs Standard et Déviant. Il s'agissait de déterminer la significativité des pics d'amplitude de l'onde de différence. L'amplitude et la latence exactes des MMNs ont ensuite été mesurées pour chaque sujet dans une fenêtre de 40 ms, centrée sur les pics. En complément, nous avons utilisé une méthode statistique appelée test non paramétrique par « partitionnement de données » dit « en cluster » (Maris et Oostenveld, 2007), pour étudier le décours temporel et la topographie exacte des évènements de négativité, les clusters étant des zones dans les représentations temporelles pour lesquelles les valeurs d'énergies diffèrent significativement entre deux conditions. Avec cette méthode, il ne s'agit plus de chercher si un point de l'espace temps-électrodes permet de rejeter l'hypothèse nulle (selon laquelle deux conditions sont équivalentes), mais de vérifier si l'on obtient un ensemble contigu de ces points, les « clusters », suffisamment grand pour ne pas être le fruit du hasard.

3 Résultats

3.1 Différences PEs Déviants – Standards et Clusters significatifs

La Fig.2 présente les signaux du grand moyennage de la différence (déviant – standard) sur Fz. Les tracés révèlent une large réponse négative, identifiée comme étant la MMN :

(i) pour la **condition non homophone** (ExpCV : Bloc[l#a] = +201 ms/-3.22 µV, Bloc[la#] = +201 ms/-2.46 µV ; ExpMot : Bloc[l#a] = +249 ms/-2.79 µV, Bloc[la#] = +247 ms/-1.71 µV),

(ii) et pour la **condition homophone** (ExpCV : Bloc[l#a] = +236 ms/-2.30 µV, Bloc[la#] = +274 ms/-1.30 µV ; ExpMot : Bloc[l#a] = +241 ms/-1.76 µV, Bloc[la#] = +182 ms/-0.86 µV).

Pour la **condition identique**, aucune MMN n'a été observée.

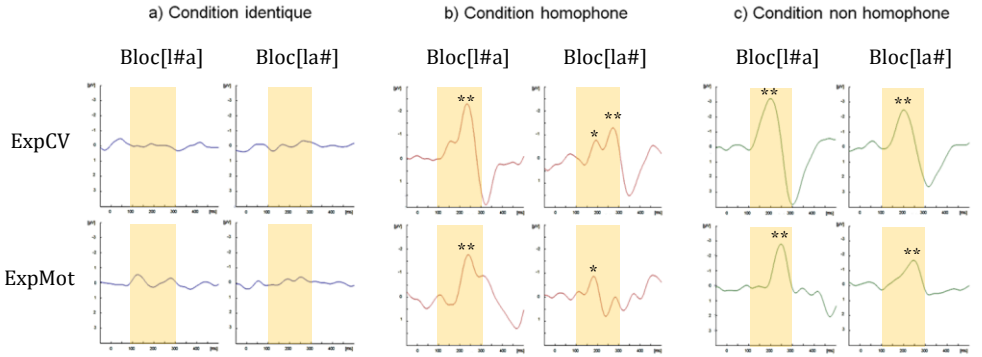

FIGURE 2 – Tracés des différences Déviant–Standard sur Fz.
La fenêtre temporelle de la MMN [100-300ms] est colorée.
Pics négatifs significatifs au rolling t-test: ** = p < .01, et, * = p < .05

Les topographies des clusters significatifs sont présentées, à leur apparition, sur la Fig.3. Toutes les négativités de notre fenêtre temporelle débutent sur quelques sites fronto-centraux, latéralisés ou non, qui se propagent ensuite sur toutes les électrodes fronto-centrales jusqu'à l'atteinte du sommet du pic négatif de la MMN.

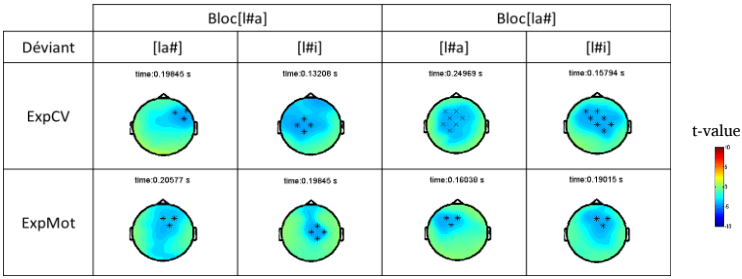

FIGURE 3 – Moment d'apparition et topographie des Clusters significatifs pour la négativité induite par les différentes déviations.

Bloc-[l#a]. Le déviant [la#] a produit une négativité fronto-centrale et **latéralisée à droite** commençant à +198 ms (fin à +241 ms) sur 3 électrodes (F4, F8, Fc6). Un cluster similaire a été observé pour le déviant *la locution* à +205 ms (fin à +269 ms) sur 3 sites (Fz, F4, Fc2). Pour la condition non-homophone, un cluster apparaît dans les 2 cas sur 4 sites centraux, latéralisé à gauche pour l'ExpCV (de +132 ms à +240 ms) et à droite pour l'ExpMot (de +198 ms à + 291 ms).

Bloc-[la#]. Pour le déviant [l#a], le cluster apparaît de manière moins localisée sur 7 électrodes fronto-centrales (de +249 ms à +289 ms), tout comme le déviant [l#i] (de +158 ms à + 249 ms). Le cluster de l'évènement négatif pour le déviant *l'allocution* apparaît **latéralisé à gauche** à +160 ms (fin à +201 ms) sur 3 électrodes frontales (F3, Fz, Fc1). Pour le déviant l'illocution, il apparaît sur 3 électrodes fronto-centrales latéralisées à droite (Fz, F4, Fc2 : de +190 ms à +278 ms).

3.2 Différences entre les MMNs

3.2.1 Condition homophone vs Condition non homophone

Dans l'ExpCV, les **MMNs observées pour la condition homophone ont été significativement plus tardives et moins négatives** que celles observées pour les conditions non homophones /li/ (+35 ms / +0.92 μV pour [la#] et +73 ms / +1.16 μV pour [l#a], ps < .001).

Les **MMNs pour *la locution* et *l'allocution* ont également été les moins négatives** (respectivement +1.03 μV et +0.85 μV par rapport aux MMNs pour *l'illocution*, ps < .001). La latence de la MMN pour le déviant *l'illocution*, bien que plus importante que celle observée pour la condition homophone *la locution* (+8 ms), n'a pas été significativement différente (t15 = 0.609, p > .1). Par contre, elle l'a été par rapport à celle du déviant *l'allocution* (+65 ms, t15 = 5.040, p < .001).

3.2.2 Bloc-[l#a] vs Bloc-[la#]

Dans l'ExpCV, la comparaison des deux conditions homophones a montré **un pic pour la MMN du déviant [l#a] significativement plus tardif et moins négatif** que celui de la MMN du déviant [la#] (+38 ms, t15 = 4.929, p < .001 ; +1.00 μV, t15 = 3.439, p < .005). Au cours de l'ExpMot, **la négativité observée pour le déviant *l'allocution* a été significativement plus précoce et moins négative** que la MMN pour *la locution* (-60 ms, t15 = -5.094, p < .001 ; +0.90 μV, t15 = 2.294, p = .037).

On peut également noter qu'aucune différence de latence n'a été observée entre les MMNs pour les déviants /li/ en fonction des blocs, et cela dans les deux expériences. En ce qui concerne les amplitudes, pour l'ExpCV comme pour l'ExpMot, les MMNs de la condition non homophone des Blocs-[l#a] ont été plus négatives que celles observées dans les Blocs-[la#] (-0.76μV pour l'ExpCV, -1.08μV pour l'ExpMot, ps < .001).

4 Discussion

La présente étude avait pour objectif de tester la robustesse des indices acoustiques fins différenciant des séquences homophones en conditions de productions de variabilité intra-locuteur, et d'examiner le décours temporel du traitement de tels indices par le

système de perception de la parole. Pour ce faire, nous nous sommes intéressés à un marqueur cortical : la MMN, obtenue à l'aide d'une version modifiée du paradigme Oddball, dans laquelle chaque stimulus provenait de productions naturelles différentes d'un même locuteur.

Nos résultats montrent clairement que les indices acoustiques qui différencient les homophones /la/ sont encodés durant le traitement de la parole et cela, malgré la variabilité des stimuli, standards comme déviants. En effet, une MMN a été obtenue pour les deux conditions homophones de l'ExpCV et de l'ExpMot. De plus, aucune MMN n'a été observée pour la condition identique (autre production d'un standard en position test) : ce qui confirme l'importance des régularités partagées par les standards variables dans la formation de la trace de mémoire sensorielle. La condition non homophone a quant à elle toujours généré des MMNs plus amples que celles observées pour les conditions homophones et dans l'ExpCV, les déviants [l#i] sont ceux qui ont été détectés le plus rapidement. Ces résultats sont en accord avec la littérature puisque *l'i* et *l'illocution* diffèrent phonologiquement des standards et des MMNs étaient par conséquent attendues. De plus, peu importe le standard (*l'a* ou *la*), les MMNs de la condition non homophone ne diffèrent pas en latence.

Un autre résultat, plus surprenant, apparait dans nos données : une asymétrie d'amplitude observée selon la nature du standard. Les blocs avec [l#a] en standard présentent en effet des MMNs plus négatives que les blocs avec [la#] en standard (en moyenne -0.9 μV) quelle que soit la nature du déviant (homophone et non homophone) et l'unité de stimulation (CV ou Mot). Cette asymétrie pourrait être sous tendue par des différences de robustesse des indices acoustiques de chaque homophone. En effet, des analyses acoustiques de nos stimuli ont montré une plus grande variabilité (durée de la syllabe ainsi que du premier formant *F1* de la voyelle) dans les productions des différents [la#] par rapport à celles des différents [l#a]. Il est ainsi envisageable que les standards [l#a], moins variables, produisent une trace plus définie permettant une détection plus efficace des déviants. Selon Näätänen et al. (2010), la trace sensorielle formée par les standards n'inclut pas seulement l'information de l'input auditif précédant mais aussi des prédictions sur les futurs évènements auditifs. En d'autres termes, la MMN serait générée lorsque les modèles prédictifs de l'environnement auditif échouent, et aurait pour fonction principale l'ajustement du modèle neuronal, permettant une meilleure description des régularités de l'environnement auditif. Ainsi, la trace formée par les standards [l#a] (*l'a* et *l'allocution*) pourrait engendrer des prédictions plus précises, ce qui augmenterait la sensibilité aux déviants et donc l'amplitude des MMNs. A l'inverse, [la#], qui correspond également à un des articles les plus fréquents de la langue française, engendre des productions naturelles bien moins stables, et ainsi des traces moins définies, ce qui a pour conséquence d'engendrer des prédictions moins précises, et donc, de rendre plus difficile la détection d'une divergence.

Une autre asymétrie a également été observée entre les 2 expériences. Dans l'ExpMot, la négativité observée pour la condition homophone *l'allocution* a en effet été très précoce (+182 ms) par rapport aux autres MMNs des deux expériences (en moyenne, +236 ms) et en particulier en comparaison avec la même condition de l'ExpCV, pour laquelle le pic d'amplitude de la MMN observée pour le déviant [l#a] est à +274 ms. Cet évènement négatif précoce pourrait avoir pour origine un mécanisme différent. Sa topographie bien

différenciée en frontale-gauche, tout comme sa précocité, semble être comparable à celle de la MMN-syntaxique de Pülvermüller et Shtyrov (2003). La trace formée par les standards *la locution* pourrait ainsi contenir une information de type grammatical, liée à la présence du déterminant [la#], dont l'absence dans le dévian*t l'allocution* produirait cette négativité précoce. Dans tous les cas, et cela est généralisable à toutes les MMNs obtenues, nos données sur la topographie d'apparition des clusters de négativité suggèrent de possible différences de générateurs de MMN. Cependant, l'interprétation de la latéralisation avec nos données de PEs ne permettent pas de conclure puisqu'aucune localisation de source n'a été effectuée.

En conclusion, la présente étude a montré que certains indices acoustiques fins sont suffisamment robustes en situation de variations intra-locuteurs pour pouvoir jouer un rôle important dans le traitement de la parole en français. Les recherches à venir viseront à identifier clairement ces indices et à clarifier les asymétries observées.

Références

BRUNELLIERE, A., DUFOUR, S., NGUYEN, N., et FRAUENFELDER, U.H. (2009). Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception, *Cognition*, 111(3), pages 390-396.

JU, M. et LUCE, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15, pages 314–318.

MARIS, E. et OOSTENVELD, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, pages 177–190.

NÄÄTÄNEN, R. et ALHO, K. (1995). Mismatch negativity – A unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80, pages 317–337.

NÄÄTÄNEN, R., ASTIKAINEN, P., RUUSUVIRTA, T., et HUOTILAINEN, M. (2010). Automatic auditory intelligence: an expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews,* 64, pages 123-136.

OOSTENVELD, R., FRIES, P., MARIS, E. et SCHOFFELEN, J.M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011, pages 1-9.

PULVERMÜLLER, F. et SHTYROV, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage*, 20, pages 159–172.

SHESTAKOVA, A., BRATTICO, E., HUOTILAINEN, M. et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport*, 13(14), pages 1813–1816.

SPINELLI, E., WELBY, P. et. SCHAEGIS, A.L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision. *Language and Cognitive Processes*, 22, pages 828–859.

SPINELLI, E., GRIMAULT, N., MEUNIER, F. et WELBY, P. (2010). An intonational cues to word segmentation in phonemically identical sequences. *Attention, Perception, & Psychophysics*, 72(3), pages 775-787.