Oscillations corticales et intelligibilité de la parole dégradée

Léo Varnet[1], *Fanny Meunier*[1], *Michel Hoen*[1]
(1) Centre de Recherche en Neurosciences de Lyon,
Inserm U1028, CNRS UMR 5292, France
leo.varnet@isc.cnrs.fr

RESUME

La méthode des potentiels évoqués a permis de caractériser différentes composantes associées au traitement de la parole. Cependant il n'existe pas aujourd'hui de marqueur cortical témoignant du succès de l'accès lexical lors de la compréhension de la parole. Le but de cette étude est donc de développer un protocole expérimental et une analyse statistique des signaux électroencéphalographiques, afin d'identifier des clusters temps-fréquence dans l'activité oscillatoire corrélant avec l'intelligibilité de stimuli paroliers. Pour mettre en évidence cet effet, nous avons présenté aux sujets des mots dégradés par noise-vocoding avant et après une courte phase d'apprentissage perceptuel. Nous avons comparé les activités oscillatoires apparaissant en réponse à des stimuli évalués comme « intelligibles » et « inintelligibles » par les participants (N = 12). Nous sommes ainsi parvenus à mettre à jour trois activités avec des topologies et des fréquences spécifiques liées au succès de l'accès lexical

ABSTRACT

Oscillatory cortical activity and intelligibility of degraded speech

Many neurocognitive aspects associated with the processing of speech were up to now studied by the analysis of event-related potentials. However, none of these cortical responses can be considered as a direct indicator of successful lexical access during speech comprehension. The aim of the present study is to develop an experimental paradigm and a statistical analysis on electrophysiological data, in order to identify time-frequency patterns in the oscillatory cortical activity that correlate with the intelligibility of degraded speech. For this purpose we used noise-vocoded speech that is very difficult to understand without prior exposure. Noise-vocoded words were presented before and after a short period of perceptual learning, and we compared the oscillatory activity following stimuli rated as "intelligible" or "unintelligible" by participants (N = 12). Results show that we were able to identify three oscillatory activities with specific topology and latency resulting from a successful lexical access.

MOTS-CLES : Intelligibilité, Noise-vocoded speech, EEG, Oscillations corticales
KEYWORDS : Intelligibility, Noise-vocoded speech, EEG, Cortical oscillatory activity

1 Introduction

Le cerveau humain est capable d'extraire le sens d'un son de parole, même dans des conditions d'audition particulièrement adverses allant de la communication téléphonique bruitée jusqu'à la discussion avec un locuteur parlant avec un accent prononcé ou dans un environnement acoustique engendrant une déformation importante. Cette faculté, dont la robustesse et la fiabilité restent jusqu'à présent inégalées par les systèmes de

Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP, pages 673–680,
Grenoble, 4 au 8 juin 2012.

reconnaissance vocale, suppose l'existence d'un processus sous-jacent, l'accès lexical, permettant l'appariement du son de parole entendu à une représentation mentale du mot reconnu. Cette représentation, stockée en mémoire au sein du lexique mental, donnerait accès à toutes les informations que le locuteur possède sur le mot en question: son orthographe, son sens et ses contraintes d'utilisation par exemple. On peut alors se demander s'il existe dans le cerveau un indicateur de la validité du mot désigné *in fine*, et s'il est possible d'identifier dans l'EEG un ou plusieurs marqueurs corticaux témoignant du succès ou non de cet accès lexical. L'analyse des réponses EEG à la présentation d'un son de parole par le biais des ERP (Event-Related Potentials) a permis de caractériser différents potentiels évoqués langagiers successifs, correspondant à des mécanismes de traitement d'informations de plus en plus haut niveau. Pourtant, aucun des potentiels tardifs mis en évidence ne reflète l'identification avec succès d'un mot entendu. Le but de cette recherche est donc de mettre en évidence un corrélat de l'intelligibilité de la parole dans les réponses EEG, en étudiant les activités cérébrales, non plus uniquement dans la dimension temporelle, comme c'est le cas pour les ERP, mais dans le domaine temps-fréquence. Cette analyse sera effectuée à l'aide des ERSP (Event-Related Spectral Perturbations), qui correspondent aux oscillations dans différentes bandes de fréquences du signal EEG. L'utilisation de cette méthode nous permettrait ainsi d'accéder aux activités non-calées en phase sur la stimulation, qui disparaissent lors du calcul des ERP.

Les précédentes études portant sur des phénomènes oscillatoires associés à une grande variété de tâches, essentiellement dans le domaine visuel, semblent indiquer que les processus de liage perceptuel s'accompagnent d'une synchronisation de réseaux de neurones dans la bande gamma (fréquences supérieures à 30 Hz), notamment pour la recherche visuelle d'un objet cohérent (Tallon-Baudry et al., 1997). Le nombre d'études se concentrant sur l'analyse des oscillations corticales pendant des tâches auditives est plus restreint, mais celles-ci ont néanmoins permis de mettre à jour des synchronisations dans la bande gamma associées à la compréhension de la parole ou à l'unification sémantique : perception d'objets auditifs cohérents (Knief et al., 2000), distinction entre mots et non-mots (Pulvermüller et al., 1996), ou accès à la mémoire pour reconstituer une parole dégradée (Hannemann et al., 2007). Par ailleurs, la synchronisation des oscillations corticales dans la bande gamma est le plus souvent envisagée comme un mécanisme servant à unifier des neurones spécialisés dans la détection d'une caractéristique particulière en un groupe de neurones représentant un certain objet perceptuel (Gray & Singer, 1989).

Parallèlement à cette augmentation de l'énergie dans la bande gamma lors de tâches sémantiques, certaines études ont mis en évidence une diminution de la puissance dans la bande alpha (8-13 Hz), dans les aires du cerveau requises pour le traitement de l'information à un instant donné. Il a été montré que l'intensité des oscillations alpha augmentait notamment avec la complexité d'une tâche de mémorisation (Jensen et al., 2002), ou avec la complexité d'une tâche de compréhension d'un son de parole dégradé (Obleser & Weisz, 2011), et ces deux équipes associent les rythmes alpha à une inhibition des oscillations gamma locales, pour permettre un accès à la mémoire à court terme ou au lexique.

Pour pouvoir identifier dans les réponses corticales du sujet à la présentation d'un son de parole une activité corrélant avec l'intelligibilité de ce son, il nous faut comparer les

réponses du sujet lorsqu'il écoute un stimulus « intelligible » et lorsqu'il écoute un stimulus « non intelligible », toutes choses égales par ailleurs. La mise en œuvre d'une telle expérience est assez délicate car elle nécessite dans un premier temps d'obtenir des signaux de parole dont on puisse faire varier l'intelligibilité sans en changer le contenu. À cette fin, nous avons utilisé des stimuli dégradés par noise-vocoding, une dégradation analogique du signal de parole qui retire une part importante des informations spectrales, tout en conservant la forme générale des enveloppes temporelles correspondant à différentes bandes de fréquence (Shannon et al., 1995). Dans le cas de notre étude, l'intérêt de l'utilisation du noise-vocoded speech tient à ce que l'intelligibilité de la parole dégradée par noise-vocoding augmente nettement après une phase d'apprentissage perceptif. Notre protocole consiste donc à comparer l'activité cérébrale durant la compréhension d'un même stimulus en noise-vocoded speech, avant et après une phase d'apprentissage. De ce fait, un stimulus inintelligible pour le sujet lors de la première phase d'écoute voit son intelligibilité notablement améliorée à la troisième phase par la période intermédiaire d'apprentissage.

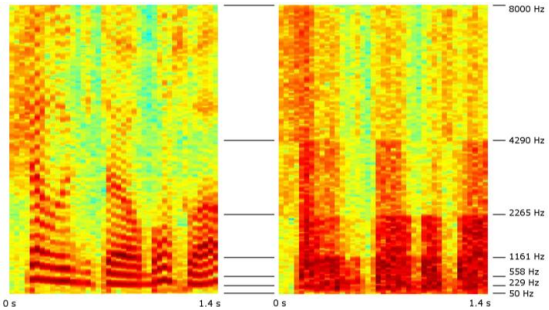

FIGURE 1 – Spectrogramme d'un son, avant (gauche) et après (droite) noise-vocoding.

2 Matériels et Méthodes

2.1 Participants

26 sujets ont pris part à cette étude. Tous étaient de langue maternelle française, droitiers, sans problèmes d'audition, ni problèmes neurologiques particuliers, et n'avaient jamais eu d'expérience préliminaire du noise-vocoded speech. Afin de ne garder que les enregistrements de très bonne qualité, nos analyses ne sont basées que sur les résultats de 12 d'entre eux (9 femmes ; âge moyen : 23,2 $\pm$ 2,4 ans), les autres étant rejetés en raison de bruits musculaires excessifs ou de problèmes techniques durant l'enregistrement.

2.2 Stimuli

Les stimuli noise-vocodés ont été générés à partir d'une liste de 400 noms français, tous dissyllabiques et débutant par une consonne. Ces stimuli ont été enregistrés par une locutrice de langue maternelle française, dans une salle insonorisée, à une fréquence

d'échantillonnage de 44.1 kHz (durée moyenne des stimuli : 480 ± 9 ms). Les enregistrements obtenus ont ensuite été coupés aux plus proches passages au zéro, puis normalisés en moyenne quadratique de l'amplitude.

Pour chacun de ces mots, une version dégradée a été créée par noise-vocoding : dans un premier temps notre algorithme filtre les enregistrements dans six bandes de fréquences espacées logarithmiquement (fréquences de coupure : 50, 229, 558, 1161, 2265, 4290, et 8000 Hz). Les enveloppes temporelles des signaux obtenus sont ensuite extraites par convolution avec une fenêtre 20-Kaiser de 64 ms, puis appliquées à un bruit limité à la bande de fréquence correspondante. Enfin, ces six bruits modulés sont additionnés pour produire le son de parole noise-vocodé.

L'ensemble de ces 400 stimuli noise-vocodés a ensuite été divisé en deux listes : l'une de 250 mots, pour les deux séquences de test, l'autre de 150 mots pour la période d'apprentissage.

2.3 Paradigme expérimental

L'expérience est divisée en trois phases : (1) test, (2) apprentissage, et (3) re-test, séparées par de courtes pauses. La durée d'une session était de 1h30 environ.

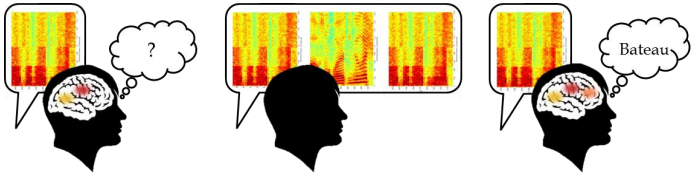

FIGURE 2 – Représentation schématique du protocole expérimental.

Les deux phases de test (phases 1 et 3) consistaient en l'écoute dans un ordre aléatoire d'une même liste de 250 mots. De cette manière, chaque stimulus était présenté deux fois au sujet, une fois avant et une fois après apprentissage. Une seconde après la présentation du mot, il était demandé au sujet par le biais d'une question affichée à l'écran d'évaluer l'intelligibilité du stimulus sur une échelle de 0 à 3 au moyen d'un système de quatre boutons (temps limité : 3 secondes), puis, si possible, de répéter le mot compris. L'essai suivant débutait après un intervalle de 0.5 seconde. Les sujets étaient encouragés à réduire leurs mouvements au minimum et si possible à ne cligner des yeux que lorsque cela leur était indiqué, afin de minimiser les artefacts oculaires. La phase d'apprentissage (phase 2) consistait en l'écoute de 150 mots noise-vocodés (différents de ceux présentés durant les phases de test et de re-test). Chacun des stimuli était suivi d'un double feedback, composé d'une version claire du mot puis à nouveau de la version dégradée (d'après les résultats obtenus par Hervais-Adelman et al., 2008). Les participants devaient écouter attentivement et appuyer sur un bouton après chaque mot. En outre, une courte séquence d'entraînement de 9 mots, semblable à la phase de test, était proposée aux participants avant le commencement de l'expérience proprement dite.

2.4 Acquisition et analyse des données

L'EEG était enregistré au moyen d'un casque à 32 électrodes actives Ag-AgCl (Biosemi, ActiveTwo) et de 8 électrodes externes placées sur le visage et les mastoïdes bilatérales pour faciliter la détection des artefacts. Les données recueillies étaient échantillonnées à 2 kHz après un filtrage passe-bas à 400 Hz.

Le traitement et l'analyse statistique des signaux ont été effectués sous FieldTrip (Oostenveld et al., 2011). Après re-référencement des électrodes à la référence moyenne, une analyse en composantes indépendantes (ACI) a été réalisée et la décomposition obtenue a été inspectée visuellement pour rejeter les composantes associées aux mouvements oculaires horizontaux et verticaux et aux artefacts cardiaques. Par ailleurs, trois électrodes trop bruitées n'ont pas été considérées pour l'analyse. L'enregistrement EEG continu a été découpé en 500 segments de 1200 ms, correspondant aux stimuli présentés durant les phases de test et de re-test uniquement, depuis 200 ms avant présentation du mot et jusqu'à 1000 ms après. Pour chaque segment, nous avons effectué une transformation en ondelettes de Morlet complexes, comme décrit dans Tallon-Baudry et al. (1997), entre 8 Hz et 140 Hz avec un pas de 1 Hz, sur toute la durée des segments, avec un pas de 50 ms. La famille d'ondelettes est définie par la formule:

$$\omega(t, f_0) = A\, e^{-i2\pi f_0 t} \exp(-t^2/2\pi\sigma_t) \;\; avec\; A = (\sigma_t \sqrt{\pi})^{-1/2} \; et \; f/\sigma_f \; = 7$$

Pour chaque signal s(t) on obtient donc une représentation temps-fréquence de l'énergie de ce signal $E(t,f) = |w(t,f) * s(t)|^2$. Pour chaque pixel temps-fréquence post-stimulus, la puissance a été ramenée à la ligne de base (entre -200 et 0 ms), puis les représentations temps-fréquence ainsi obtenues pour chaque essai ont été ensuite moyennées indépendamment pour chaque condition pour obtenir les ERSP, qui reflètent donc les variations moyennes dans le spectre de puissance de l'activité cérébrale par rapport à son activité basale.

L'analyse statistique a été effectuée par un test statistique en mesures répétées sur les mots considérés par les participants comme plus intelligibles dans la troisième phase que dans la première (c'est-à-dire dont la note d'intelligibilité attribuée par le sujet est plus importante après apprentissage qu'avant). Les représentations temps-fréquence correspondant à ces mots ont été intégrées à un test non paramétrique basé sur les clusters, comme décrit par Maris et Oostenveld (2007). Cette procédure effectue un test-t apparié sur l'ensemble des couples, comparant la puissance au niveau de chaque point temps-fréquence avant et après apprentissage. Le résultat est ensuite corrigé pour les comparaisons multiples par un test non-paramétrique basé sur les clusters. Il ne s'agit plus seulement de chercher si un point de l'espace temps-fréquence permet de rejeter l'hypothèse nulle mais de vérifier si l'on obtient un ensemble contigu de ces points, suffisamment grand pour ne pas être le fruit du hasard. La statistique permettant de décrire ces clusters est T_{sum}, la somme des t-values à l'intérieur du cluster. Comme nous ne connaissons pas la loi de probabilité suivie par T_{sum}, nous devons re-répartir aléatoirement nos données entre les deux conditions un nombre de fois suffisamment grand (500 itérations dans notre cas) pour obtenir une estimation de la distribution de T_{sum} correspondant au cas de l'hypothèse nulle. Il est alors possible de comparer la valeur de T_{sum} obtenue expérimentalement à cette distribution calculée pour décider s'il est pertinent de rejeter l'hypothèse nulle, avec un taux d'erreur donné.

3 Résultats

3.1 Résultats comportementaux

Une ANOVA à mesures répétées réalisée sur les taux de réponses correctes et sur les notes d'intelligibilité attribuées par les sujet à chaque stimulus entre les conditions Avant apprentissage puis Après apprentissage nous indique que l'évolution de l'intelligibilité a, comme attendu, un effet significatif sur les deux variables ($p < .001$). Ceci valide ainsi notre protocole expérimental, et nous autorise à chercher dans un second temps un corrélat électrophysiologique de cette augmentation de l'intelligibilité.

3.2 Résultats électrophysiologiques

L'analyse statistique de nos données EEG a abouti au résultat suivant, présenté sur la figure 3

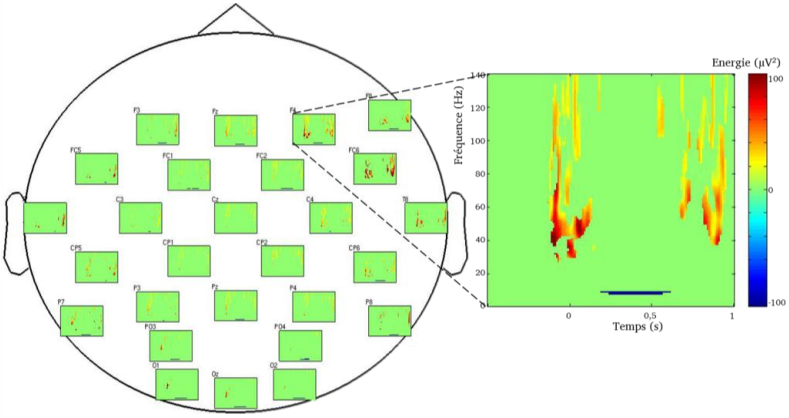

FIGURE 3 – Représentations des différences significatives d'énergie (ou « clusters ») entre les conditions Avant et Après apprentissage sur l'ensemble du scalp et au niveau de F4

Au niveau de chaque électrode est représentée la différence de puissance moyenne entre les conditions Avant et Après apprentissage. On ne conserve ensuite que les clusters significatifs obtenus par l'analyse statistique, le reste de la représentation étant mis arbitrairement à la valeur 0. On constate la présence de trois clusters significatifs ($p < .001$). Le plus important apparaît dans la bande gamma, pour des latences entre 500 ms et 1 s, et il est distribué sur l'ensemble du scalp mais essentiellement localisé dans les zones temporales bilatérales. L'autre cluster dans la bande gamma apparaît 120 ms avant l'apparition du stimulus et se termine 120 ms après. Il présente une distribution topologique similaire. Ce sont deux clusters positifs (en rouge sur la figure 3), c'est-à-dire qu'ils correspondent à une augmentation de la puissance dans la condition Après apprentissage par rapport à la condition Avant apprentissage. Enfin, on constate également la présence d'un troisième cluster, négatif celui-ci, situé dans la bande alpha

pour des latences entre 200 ms et 700 ms (en bleu sur la figure 3). Il est majoritairement distribué sur les électrodes occipitales, mais aussi au niveau de l'aire frontale.

4 Discussion

Notre recherche nous a donc permis d'identifier trois clusters liés à l'intelligibilité des stimuli langagiers présentés : un premier cluster positif précoce dans la bande gamma, un deuxième, négatif dans la bande alpha et enfin un dernier cluster positif plus tardif dans la bande gamma.

La diminution de l'activité alpha suivie d'une augmentation de l'activité gamma est une observation classique des modulations de l'activité oscillatoire dans de nombreuses tâches (motrices, mémorielles, attentionnelles…) et notamment lors de tâches de compréhensions de la parole (Shahin et al., 2009 ; Obleser & Weisz, 2011). Les deux clusters les plus tardifs de notre étude semblent correspondre à ce couple désynchronisation/synchronisation. Le cluster gamma tardif serait une marque de l'accès lexical et/ou sémantique. En effet, plusieurs études antérieures montrent une association entre l'augmentation de la puissance dans la bande gamma et les processus de compréhension de la parole (Hannemann et al., 2007 ; Shahin et al., 2009). Par ailleurs, la diminution significative de l'amplitude des oscillations dans la bande alpha, observée lors de l'écoute d'un stimulus intelligible, semble elle aussi refléter une augmentation dans l'intensité des traitements de la parole (voir Obleser & Weisz, 2011). Plus précisément, les oscillations dans la bande alpha distribuées sur l'ensemble du scalp pourraient être envisagées comme un mécanisme d'inhibition des oscillations haute-fréquence plus locales mentionnées ci-dessus, ce qui, dans le cas de stimuli inintelligibles, permettrait au système de ne pas considérer les éventuelles activations dans les aires liées au langage. Cette interprétation est soutenue par certaines études effectuées sur les processus d'inhibition qui montrent une corrélation entre l'augmentation de la puissance dans la bande alpha et la rétention de la mémoire de travail (voir par exemple Jensen et al., 2002). Cette première conclusion doit être cependant relativisée par la présence d'un facteur confondu, l'apprentissage, qui sera étudié indépendamment lors d'une prochaine étude.

Le premier cluster apparaissant dans la bande gamma reste le plus surprenant et le plus délicat à interpréter. En effet, il apparaît dès 120 ms avant la présentation du stimulus, alors même que notre analyse ne prend en compte que des clusters liés à l'intelligibilité du stimulus présenté. Comment cette dernière peut-elle influer l'activité cérébrale du sujet alors que le mot n'a pas encore été présenté ? On doit alors envisager que c'est à l'inverse, la présence de cette activité qui permet une meilleure intelligibilité du stimulus. Il s'agirait dès lors d'un cluster d'anticipation et de préparation à la perception.

5 Conclusion

Cette étude nous a permis de mettre en évidence un marqueur de l'intelligibilité cérébrale, c'est-à-dire une activité corrélant avec une amélioration de la qualité de la compréhension du stimulus par le sujet. Notre résultat offre à cette impression subjective, inaccessible à tout autre que celui qui la perçoit, un substrat matériel et une objectivation phénoménologique observable par des tiers. Les recherches futures seront

dédiées à la distinction plus précise des effets d'intelligibilité *per se* de ceux de l'apprentissage perceptif.

Remerciements

Ce projet de recherche est soutenu par un financement du Conseil Européen de la Recherche (ERC), starting-grant attribuée au second auteur (SpiN Project, ERC n°209234).

Références

GRAY, M. et SINGER, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Neurobiology, 86,* pages 1698-1702.

HANNEMANN, R., OBLESER, J., EULITZ, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research, 1153,* pages 134-143.

HERVAIS-ADELMAN, A., DAVIS, M., JOHNSRUDE, I., CARLYON, R. (2008). Perceptual Learning of Noise Vocoded Words: Effects of Feedback and Lexicality. *Journal of Experimental Psychology, 34,* pages 460–474.

JENSEN, O., GELFAND, J., KOUNIOS, J., LISMAN, J.E. (2002). Oscillations in the alpha band increase with memory load during retention in a short-term memory task. *Cerebral Cortex, 12,* pages 877-882.

KNIEF, A., SCHULTE, M., BERTRAND, O., PANTEV, C. (2000). The perception of coherent and non-coherent auditory objects: a signature in gamma frequency band. *Hearing research, 145,* pages 161-168.

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

OBLESER, J., et WEISZ, N. (2011). Suppressed alpha oscillations predict intelligibility of speech and its acoustic Details. *Cerebral Cortex, In press.*

OOSTENVELD, R., FRIES, P., MARIS, E., SCHOFFELEN, JM. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, Volume 2011*

PULVERMÜLLER, F., EULITZ, C., PANTEV, C., MOHR, B., FEIGE, B., LUTZENBERGER, W., ELBERT, T., BIRBAUMER, N. (1996). High-frequency cortical responses reflect lexical processing: an MEG study. *Electroencephalography and clinical Neurophysiology, 98,* pages 76-85.

SHAHIN, A.J., PICTON, T.W., MILLER, L.M. (2009). Brain oscillations during semantic evaluation of speech. *Brain Cognition, 70,* pages 259-266.

SHANNON, R.V., ZENG, F.-G., WYGONSKI, J., KAMATH, V., EKELID, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* pages 303-304.

TALLON-BAUDRY, C., BERTRAND, O., PERONNET, F., PERNIER, J. (1997) Oscillatory γ-Band Activity Induced by a Visual Search Task in Humans. Journal of Neuroscience, 18, pages 4244-4254.