Mesure d'intelligibilité de segments de parole à l'envers en français : une étude électrophysiologique.

Hoen, M., Maurin, A.E., Dodane, C, & Meunier, F.

Laboratoire Dynamique du Langage
Institut des Sciences de l'Homme, 14, avenue Berthelot – 69363 Lyon Cedex 07, France
Tél.: +33 (0)4 72 72 64 12 - Fax: +33 (0)4 72 72 65 90
Mél: fanny.meunier@univ-lyon2.fr; aude.maurin@etu.univ-lyon2.fr - http://www.ddl.ish-lyon.cnrs.fr

Institut des Sciences Cognitives
67 bd Pinel – 69675 Bron Cedex, France
Tél.: +33 (0)4 37 91 12 12 - Fax: +33 (0)4 37 91 12 10
Mél: hoen@isc.cnrs.fr; dodane@isc.cnrs.fr- http://www.isc.cnrs.fr

ABSTRACT

We present results from an experiment studying the cognitive ability of reversed speech restoration in French. Our results show high performance loss after reversions of at least one syllable or more. ERP recordings show the presence of an MMN, related to the processing of physical distortions in auditory stimuli. A second wave, labelled N600 or 'restoration negativity', displayed a left anterior spatial distribution and was clearly associated to the cognitive restoration process. Results are compared to recent imagery studies showing left frontal brain regions implicated in the restoration of various types of degraded auditory stimuli.

1. INTRODUCTION

La perception du langage parlé est une activité menée quotidiennement, qui s'effectue de façon rapide et automatique malgré des processus cognitifs sous-jacents complexes [5]. Il a été établi que pour associer un 'sens' à une 'forme auditive' l'auditeur ne se base pas uniquement sur une analyse directe du signal sonore mais qu'il utilise également sa capacité à générer des inférences [2]. L'intelligibilité d'un signal de parole va donc dépendre de l'interaction entre la clarté du signal émis et la capacité cérébrale à traiter ce signal. Cette capacité implique notamment des aptitudes cognitives de correction d'erreurs dues à des distorsions du signal et de restauration des éléments auditifs non-perçus. Différentes études ([11] et [3] par exemple) ont montré que le langage parlé reste dans une certaine mesure intelligible malgré l'application de détériorations acoustiques au signal : il existe une capacité cognitive de restauration du signal de parole.

Différentes sortes de détériorations acoustiques ont été étudiées ; parmi elles, l'inversion temporelle de la parole a été qualifiée comme '*la forme la plus drastique*' [9]. Le langage à l'envers, tout en gardant certaines propriétés acoustiques du langage à l'endroit, comme la répartition fréquentielle des sons, leur amplitude et certaines caractéristiques rythmiques, viole beaucoup de propriétés phonologiques segmentales ou supra-segmentales qui sont universellement observées dans le langage humain [10]. Cependant le système cognitif humain reste capable dans certaines conditions de restaurer la parole inversée. Différentes expériences ont montré que l'intelligibilité du signal dépend de la longueur et de la fréquence des fenêtres d'inversion [9] [3] [7]. Notamment il a été établi qu'une restauration quasi parfaite était possible si l'inversion était effectuée sur des fenêtres temporelles inférieures à 50 ms.

Le but de l'expérience présentée est de poursuivre ces travaux, afin de quantifier les capacités du système cognitif à récupérer l'information lexicale de phrases contenant un mot ayant subi une inversion temporelle, grâce à une mesure comportementale. Des Potentiels Evoqués ont également été enregistrés afin d'identifier les corrélats électrophysiologiques de ces processus. Nous voulions établir si la latence et/ou l'amplitude de la Mismatch Negativity (MMN), marqueur associé à la perception de stimuli auditifs déviants, varie en fonction du type d'inversion temporelle appliquée au signal. Nous cherchions également à voir si d'autres marqueurs évoqués reflètent la perte d'intelligibilité ou la phase de reconstruction cognitive.

2. EXPERIENCE

Le principe de l'expérimentation est de soumettre des sujets à l'écoute d'une série de phrases contenant un mot plus ou moins modifié et de leur demander de répéter la phrase qu'il leur semble avoir entendu.

2.1 Méthode

Matériel

Un corpus de 160 phrases est constitué. Les phrases comportent de 7 à 9 mots (m = 7,71 ; ET = 0,67). Les phrases ont toutes la même structure : sujet-verbe-complément-complément : *Exemple : Le baron offre des bijoux à une princesse.* Chaque phrase contient 3 noms bisyllabiques. Elles sont prononcées par une seule locutrice. L'inversion temporelle du signal de parole se fait par l'intermédiaire du logiciel Matlab et

consiste en un retournement du signal sonore au long de son axe temporel, sur le segment phonémique considéré. L'inversion du signal de parole est appliquée sur le mot cible qui est l'un des 3 noms de la phrase, ceci permettant de d'assurer de la non prédictibilité de la position d'apparition de l'inversion. Les fenêtres d'inversion sont définies selon un critère utilisant une unité linguistique, la syllabe. Quatre conditions expérimentales sont testées : une condition 0 (C0) ou contrôle, dans laquelle le mot cible reste intact, une condition 0.5 (C0.5), dans laquelle la moitié de la première syllabe est inversée, la condition 1 (C1), dans laquelle la première syllabe est inversée et la condition 1.5 (C1.5) dans laquelle une syllabe et demie est inversée. Pour chaque sujet, chacune des 160 phrases n'apparaît qu'une fois et dans une seule condition donnée (C0, C0.5, C1 ou C1.5).

Déroulement de l'expérience

Les sujets sont placés à 30 cm d'un écran vidéo sur lequel sont affichés les consignes marquant le déroulement temporel de l'expérience. Les phrases sont émises par des enceintes standard.

La tâche demandée aux sujets est d'écouter chacune des 160 phrases puis de les répéter à voix haute, immédiatement après l'écoute de la phrase perçue. Chacune des phrases ne peut être écoutée qu'une seule fois. Les sujets doivent essayer de dire la phrase qu'il leur semble avoir entendu en utilisant des mots français. La restitution correcte ou non du mot cible est notée. Un enregistrement EEG continu est réalisé en simultané. La durée totale de l'expérience est d'environ 2 heures.

Recueil des données EEG

Les courants de surface sont collectés grâce à 64 électrodes souples (Ag/AgCl), à électrolyte liquide (KCl), montées sur un filet à structure géodésique (Geodesic Sensor Net^{TM}). Le signal de scalp est amplifié par un amplificateur à haute impédance (200MΩ, Net AmpsTM, Electrical Geodesics INC.), sur une bande passante de 0.1 à 200Hz. La fréquence d'échantillonnage est de 500Hz avec des impédances d'électrodes maintenues inférieures à 50kΩ. Durant l'enregistrement, une 65eme électrode, placée au niveau du Vertex, défini comme l'intersection médiale de l'axe antéro-postérieur (Nasion - Inion) avec l'axe coronal (point auriculaire gauche – point auriculaire droit, Jasper, 1958), sert d'électrode de référence.

Sujets

Vingt trois sujets volontaires (11 f, 12 h), âgés de 18 à 37 ans, de langue maternelle française, et n'ayant jamais connu aucun troubles auditifs, du langage, ou neurologiques, ont participé à l'expérience. Tous les participants étaient naïfs par rapport au but de l'étude et ont été indemnisés pour leur participation.

2.2 Résultats

Analyse des mots cibles reconnus :

Une analyse ANOVA a été réalisée à l'aide du logiciel SPSS, en considérant comme variable aléatoire d'une part les sujets (F1) et d'autre part les items (F2). La variable dépendante est le pourcentage de rappel correct des mots cibles expérimentaux.

Table 1. Pourcentage de rappel correct du mot cible entendu dans sa phrase de contexte.

Type d'inversion	C0	C0,5	C1	C1.5
Taux de reconnaissance ET	99.7 (1.1)	99.6 (1)	87.5 (4.8)	53.1 (8.5)

Globalement on observe une différence significative entre les conditions expérimentales par sujets et par items ($F1(3,66)= 526.09$; $p<.0001$) ; $F2(3,156)= 59.29$; $p<.0001$). Des comparaisons spécifiques montrent qu'il n'y a pas de différence significative entre les conditions C0 et C0.5 ($F1<1$; n.s ; $F2<1$; n.s.), alors qu'il en existe une entre C0.5 et C1 ($F1(1,22)= 165.28$; $p<.0001$; $F2(1,78)= 15.31$; $p<.0002$) et entre C1 et C1.5 ($F1(1,22)= 420.29$; $p<.0001$; $F2(1,78)= 36.53$; $p<.0001$).

Analyse des enregistrements de potentiels évoqués

Traitement des données

Pour chaque sujet, l'enregistrement brut est segmenté sur une fenêtre temporelle allant de 100 ms avant, à 900 ms après l'occurrence du mot cible. Les segments d'EEG sont alors moyennés entre-eux, pour chaque condition. Durant cette phase de moyennage, un rejet automatique d'artefacts est appliqué, supprimant les artefacts dus aux mouvements oculaires, aux sauts transitoires de potentiel et aux électrodes bruitées. Les segments restant sont alors exportés sous format numérique vers le logiciel BESA 2000 (MEGIS Software GmbH, Munich, Allemagne), dans sa version 4.2.24, où l'efficacité du rejet automatique est vérifiée par inspection visuelle des segments. Par ailleurs, un enregistrement est complètement rejeté des étapes ultérieures de l'analyse s'il contient plus de 10% d'essais rejetés ou plus de 10% d'électrodes artéfactées (6 électrodes/64). Pour cette expérience, les enregistrements de 8 sujets ont été rejetés, l'analyse a donc été faite sur les 16 sujets restants (8h/8f).

Moyennage et grand-moyennage

Les segments conservés sont normalisés en prenant les 100 ms précédant l'apparition du stimulus pour ligne de base. Les tracés sont re-référencés par rapport à une référence moyenne virtuelle. Enfin, les enregistrements sont filtrés par application d'un filtre digital passe-bas à 30 Hz. Les enregistrements sont alors moyennés entre eux, pour chaque condition et séparément pour chaque sujet avant analyse statistique, l'ensemble des enregistrements des différents sujets sont également

moyennés entre-eux, afin d'être visualisés sous forme 'Grand-Moyen'.

Analyse statistique

Une analyse de variance en plan de mesures répétées (ANOVA-RM, $\alpha = 0.05$) est effectuée, en prenant la valeur moyenne de l'amplitude du voltage de surface, calculée dans une fenêtre temporelle et exprimée en microvolts, pour variable dépendante. Lorsqu'un effet possède plus d'un degré de liberté au numérateur, la correction de Greenhouse-Geisser/Huynh Feldt est appliquée. Une première analyse comporte les facteurs : Fenêtre Temporelle (2 : Précoce/Tardive), Condition (4 durées de la fenêtre d'inversion : 0, 0.5, 1 ou 1.5 syllabes) et Electrodes (20 électrodes réparties de façon homogène à la surface du scalp et correspondant aux électrodes du système 10-20 : PO4, C1, AFz, F3, F5, FC3, Cp1, T9, Tp7, P3, Pz, POz, Cp2, P6, TP8, T8, FC4, C2, F8, F4). Afin de préciser les observations de cette première analyse et étant donné que les marqueurs évoqués observés possèdent une distribution spatiale focale et asymétrique à la surface du scalp, une seconde ANOVA incluant un facteur Domaine Spatial a été réalisée (5, incluant chacun 5 électrodes : Fronto-Central (FC : AF4, AFz, AF3, FP2 et FP1), Frontal Gauche (FG : F1, F3, FC1, FC3 et AF7), Frontal Droit (FD : AF8, Fz, FC4, FC4 et F4), Central (C : C1, Cp1, Cpz, Cp2, C2) et Postérieur (P : P3, P6, PO3, PO4 et POz). Les fenêtres temporelles sont définies par rapport à la latence du pic d'amplitude des effets principaux et s'étendent de 100 à 300 ms après apparition du mot cible pour l'onde MMN et de 400 à 750 ms pour l'onde frontale.

Marqueurs Evoqués :

L'introduction d'une fenêtre d'inversion non-nulle dans le mot cible fait apparaître 2 modifications principales dans le tracé du potentiel évoqué par l'audition de ce mot (voir Figure 1). Une première onde focale est observable dans une fenêtre temporelle précoce, cette onde possède une polarité négative, une distribution spatiale centro-pariétale, un pic d'amplitude proche de Cz ayant une latence entre 180 et 200 ms et une ligne d'inversion de ses valeurs de potentiel au niveau des électrodes mastoïdes. Cette onde peut être classifiée comme une onde de type MMN [8]. Suivant la MMN, une seconde onde, possédant également une polarité négative apparaît. Sa distribution spatiale est globalement frontale, avec une légère asymétrie en faveur de l'hémi-scalp gauche et un pic d'amplitude mesuré au niveau de l'électrode F7 ayant une latence approximative de 600 ms.

La première ANOVA, portant sur les facteurs Fenêtre Temporelle (FT :2), Condition (C : 4) et Electrodes (E : 20), révèle un effet principal non significatif du facteur FT ($F(1)= 1.46$; n.s.), les valeurs moyennes de potentiel étant globalement négatives dans ces deux fenêtres temporelles. En revanche, cette analyse montre un effet principal significatif du facteur Condition ($F(3,15)= 5.47$; $p<.05$), les valeurs étant globalement plus négatives pour C1 et C1.5 que pour les fenêtres d'inversion plus courtes. L'effet du facteur Electrodes ($F(19,285)= 16.96$; $p<.05$) était également significatif. L'interaction significative FT*C : ($F(3,285)= 3$; $p<.05$), suggère que l'effet de la condition ne module pas les deux ondes de la même manière. Les autres interactions de second et troisième niveau sont également significatives.

Une seconde ANOVA, incluant le facteur Domaine Spatial (5) a été réalisée en considérant uniquement la fenêtre temporelle tardive (450 à 700 ms), afin de préciser ces effets. Un test post-hoc de type LSD, effectué sur l'interaction de second ordre significative Condition*Domaine Spatial révèle que l'onde négative frontale gauche est plus ample pour les conditions d'inversion à fenêtres longues, 1 et 1.5 syllabes et ce dans les domaines spatiaux FC et en FG, confirmant la distribution spatiale frontale-gauche de cette onde.

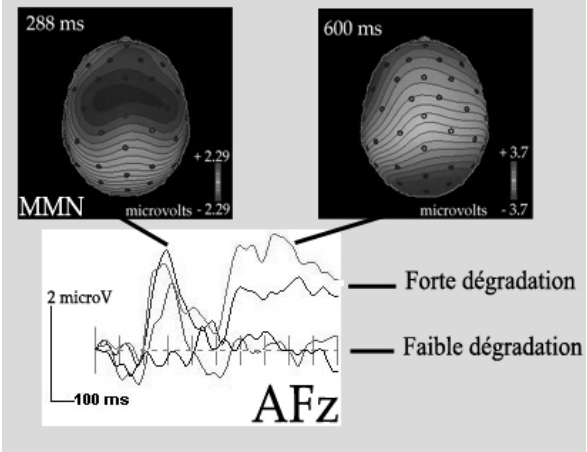

Figure 1 : Bas : électrode frontale (AFz), valeurs de voltage de scalp en fonction de la condition de dégradation. Faible dégradation : C0 (noir) et C0.5 (rouge). Forte dégradation : C1 (bleu) et C1.5 (rose). Haut gauche : interpolation 2D des courants de surface à 288 ms, onde MMN. Haut droite : idem à 600 ms, onde N600.

La présence d'une portion de parole inversée provoque en EEG l'apparition d'une onde MMN dépendant de la présence ou de l'absence d'une inversion sonore, pas de son importance temporelle. Cette onde MMN est suivie d'un marqueur frontal gauche, à notre connaissance non rapporté dans une autre expérience à ce jour, que nous appellerons donc onde de reconstruction ou N600. Cette onde de reconstruction apparaît modulée par la taille de la fenêtre d'inversion.

2.3 Discussion

Les résultats des données comportementales montrent une parfaite restauration du signal lorsque l'inversion ne touche que la moitié de la première syllabe (C0.5) en revanche une baisse des performances est observée lorsque la dégradation est sur la 1ère syllabe (C1) -87,5 % de reconnaissance- et celle-ci baisse encore (53,1 %) lorsque lorsqu'une syllabe et demie est inversée. Ces

résultats sont compatibles avec ceux observés dans les expériences précédentes.

Les résultats obtenus pour les enregistrements de potentiels évoqués, montrent quant à eux, que les trois types d'inversion provoquent l'apparition du marqueur des stimuli déviants, la MMN. D'autre part, l'inversion sur des segments phonémiques dont la longueur est égale à une syllabe (C1) ou à une syllabe et demie (C1.5) provoque l'apparition d'une onde suivant celle de la MMN, une N600, qui pourrait être une onde liée à la reconstruction cognitive.

Une recherche avec laquelle il semble intéressant de confronter nos résultats est celle de Davis et Johnsrude [1]. Ces auteurs ont utilisé dans leur expérience d'IRMf 3 types de distorsion, avec chacune 3 niveaux d'intelligibilité. Ce choix visait à identifier les régions cérébrales dont le signal BOLD variait avec l'intelligibilité du stimulus, ceci indépendamment du type de distorsion appliqué. Leurs résultats mettent en évidence l'existence d'un réseau d'aires corticales dont le niveau d'activation est insensible aux formes acoustiques des phrases détériorées. Ces aires sont situées à l'intérieur du gyrus temporal, de l'hippocampe et du gyrus frontal inférieur gauche. Au sein de ce réseau, certaines aires sont particulièrement actives dans les conditions complètement inintelligibles et en conséquence doivent être impliquées dans le phénomène de compensation de la détérioration.

Les résultats que nous avons obtenus en potentiels évoqués paraissent compatibles avec ces observations. En effet, il semble probable que l'onde de reconstruction (N600) observée dans notre expérience soit générée au niveau des régions du cortex préfrontal gauche, sans que nos données ne permettent pour le moment de l'affirmer de façon certaine.

3. CONCLUSION

Dans cet article, nous avons présenté une expérience sur la capacité cognitive à restaurer de la parole inversée en français. Nos résultats montrent une forte dégradation des taux de compréhension lorsque la fenêtre de réversion est égale à une syllabe. Les PEs révèlent l'apparition d'une MMN, liée au traitement de la distorsion physique du stimulus. Une seconde onde, frontale gauche, semble corrélée au processus de reconstruction cognitive du message sonore. Des expériences complémentaires seront menées afin de clarifier les processus cognitifs sous-jacents à cette tâche de restauration de parole inversée.

4. NOTES DES AUTEURS ET REMERCIEMENTS

Nous remercions Lionel Granjon pour son aide. Cette expérience a été réalisée grâce à une ACI – MSH 2001 attribuée à Fanny Meunier.

BIBLIOGRAPHIE

[1] M.H. Davis and I.S. Johnsrude. Hierarchical Processing in language comprehension, *Behavioral / Systems Neuroscience*, p. 35, 2003.

[2] D.P.W. Ellis. Using knowledge to organize sound : The prediction-driven approach to computational auditory scene analysis, and its application to speech / nonspeech mixtures, *Speech Communications*, *Special issue on Computational Auditory Scene Analysis*, M. Cooke and H. Okuno, guest editors, 1998.

[3] S. Greenberg and T. Arai. The relation between speech intelligibility and the complex modulation spectrum, P*roceedings of the 7th Eurospeech Conference on Speech Communication and Technology,* pp. 473-476, 2001.

[4] H.H. Jasper. The Ten-Twenty Electrode System of the International Federation. *Electro-encephalography and Clinical Neurophysiology.* vol. 10, pp. 371-375, 1958.

[5] R. Kolinsky, J. Morais and J. Segui. *La reconnaissance des mots dans les différentes modalités sensorielles, études de psycholinguistique cognitive.* Puf, 1991.

[6] J. Melher, P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini and C. Amiel-Tison. A precursor of language acquisition in young infants, *Cognition*, 29, 143-178, 1988.

[7] F. Meunier, T. Cenier, M. Barkat and I. Magrin-Chagnolleau. Mesure d'intelligibilité de segments de parole à l'envers en français, *XXIVèmes Journées d'Etude sur la parole*, Nancy, 2002.

[8] R. Näätänen, A.W.K. Gaillard and S. Mäntysalo. Early selective-attention effect on evoked potential reinterpreted, *Acta Psychologia*, 42, 313-329, 1978.

[9] K. Saberi and D.R. Perrott, (1999). Cognitive restoration of reverse speech, *Nature,* 398: 760

[10] J. Vaissière. *In Prosody : Models and Measurements*, A. Culter, Ladd, D. R., Eds (Springer-Verlag, New-York, pp.53-66, 1983.

[11] R. M. Warren. Perceptual restoration of missing speech sounds, *Science,* 167, 392-393, 1970.