



HAL
open science

Masques acoustiques et masques linguistiques de différentes langues sur la reconnaissance de mots en français

Aurore Gautreau, Michel Hoen, Fanny Meunier

► **To cite this version:**

Aurore Gautreau, Michel Hoen, Fanny Meunier. Masques acoustiques et masques linguistiques de différentes langues sur la reconnaissance de mots en français. Conférence conjointe JEP-TALN-RECITAL, Jun 2012, Grenoble, France. pp.755-762. hal-00941158

HAL Id: hal-00941158

<https://hal.science/hal-00941158v1>

Submitted on 3 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Masques acoustiques et masques linguistiques de différentes langues sur la reconnaissance de mots en français

Aurore Gautreau¹, Michel Hoen¹, Fanny Meunier¹

(1) Centre de Recherche en Neurosciences de Lyon,
INSERM U1028 - CNRS UMR5292, France

aurore.gautreau@gmail.com

RESUME

Nos recherches visent à explorer les interférences linguistiques qui ont lieu dans la situation de compréhension de la parole dans la parole. Pour cela, l'intensité, la nature et la langue du bruit de fond concurrent ont été manipulées. Des participants français ont réalisé une tâche de décision lexicale sur des items cibles français insérés à 0 dB et à -5 dB dans des masqueurs de parole ou des masqueurs de bruit fluctuant générés en français, gaélique irlandais et italien. A -5 dB, les résultats ont montré que le français et l'italien ont davantage masqué la parole cible que l'irlandais. La comparaison des performances obtenues entre les deux types de masqueurs (paroliers versus bruit fluctuant) a révélé que pour le français, la dégradation produite sur la parole cible était à la fois de nature acoustique et linguistique alors qu'elle n'a été que de nature acoustique pour l'italien et l'irlandais.

ABSTRACT

Acoustical and linguistic masking effects of different languages on the comprehension of French words

The goal of our research is to explore linguistic interferences which occur during speech-in-speech comprehension. Intensity, nature and language of the background noise were manipulated. Native French participants had to realize a lexical decision task on French target items inserted at 0 dB or -5 dB in background speech or fluctuating noise produced in French, Irish and Italian. At -5 dB, results showed that French and Italian had a stronger masking effect on target speech than Irish. A comparison of performances obtained with background speech and fluctuating noise revealed that for French, target speech was masked by acoustic and linguistic information whereas the degradation from Italian and Irish was only acoustic.

MOTS-CLES : masqueur parolier, bruit fluctuant, compréhension de la parole, interlangue.

KEYWORDS : babble noise, fluctuating noise, speech comprehension, interlanguage.

1 Introduction

Comprendre la parole en présence d'un bruit de fond est une tâche complexe, d'autant plus lorsque le signal concurrent contient de la parole plutôt qu'un bruit dépourvu d'informations linguistiques. En effet, en plus d'appliquer un effet de masque énergétique sur la parole cible (superposition au moins partielle des informations spectro temporelles des deux sources sonores), la parole concurrente est également responsable d'un effet de

masque informationnel (Bronkhorst, 2000 ; Brungart, 2001). Différents types d'informations sont en compétition, notamment des informations de nature linguistique (phonétiques, lexicales, sémantiques) ou psycholinguistique (genre de la voix, F0...). Notre étude a pour objectif d'explorer les interférences linguistiques qui ont lieu dans la situation particulière de compréhension de la parole dans la parole.

Afin de déterminer la nature des informations linguistiques qui entrent en compétition, des études récentes ont manipulé la langue de la parole concurrente. Elles ont montré que la langue de la parole du bruit de fond pouvait affecter la compréhension de la parole cible. Par exemple, Rhebergen et al. (2005) ont observé une moins bonne identification de phrases cibles allemandes lorsque les masqueurs étaient produits en allemand plutôt qu'en suédois. Van Engen et Bradlow (2007) ont obtenu des performances plus faibles pour identifier des phrases cibles anglaises avec des masqueurs produits en anglais plutôt qu'en mandarin. Ces études ont montré que l'effet de masque du bruit de fond était plus important lorsque son contenu linguistique était identique à celui de la parole cible, autrement dit lorsque les deux signaux de parole en compétition étaient produits dans la même langue. Cependant, ces expériences ne permettent pas de savoir si l'effet observé provient du fait qu'on utilise une langue différente de la parole cible ou si des caractéristiques linguistiques des langues manipulées vont dégrader différemment la parole cible. Afin d'étudier plus avant cette question, nos expériences ont comparé l'effet de masque de différentes langues du monde sur des mots cibles français.

1.1 La présente étude

Dans cette étude, les items cibles étaient toujours produits en français, alors que les masqueurs ont été générés en français, italien et gaélique irlandais. Il s'agissait d'explorer des différences d'effets de masque dont l'origine pourrait provenir de caractéristiques linguistiques différentes entre ces langues. Pour sélectionner ces langues, de nombreux critères étaient disponibles. Dans une première approche, nous nous sommes intéressés à l'inventaire phonologique qui correspond à la description des différents phonèmes d'une langue.

Le français compte 37 phonèmes soit 16 voyelles et 21 consonnes (Maddieson et al., 2011). Sur la base de cette description nous avons pu établir que l'italien possède 60% de ses phonèmes qui sont identiques à ceux du français. Pour cela, l'italien sera considéré comme proche du français (langue cible) contrairement au gaélique irlandais qui partage seulement 18% de ses phonèmes avec le français. Deux expériences ont été menées à des RSB (Rapport Signal sur Bruit) différents : 0 dB et -5 dB. Pour chaque expérience, l'effet de masque du français sur la compréhension des items cibles français a été comparé aux effets de masque d'une langue éloignée (irlandais) et d'une langue proche du français (italien). Tout d'abord, il s'agissait de déterminer si l'identification des items cibles français serait plus affectée par les masqueurs français (langue identique à celle de la parole cible) que par des masqueurs produits dans une langue non comprise par les participants. Un second objectif était d'observer si les masqueurs produits dans une langue proche du français dégraderaient davantage la parole cible que les masqueurs produits dans une langue éloignée du français. Les participants devaient réaliser une tâche de décision lexicale sur des items cibles français insérés dans les masqueurs.

2 Expériences 1 et 2

L'expérience 1, réalisée à 0 dB, et l'Expérience 2 à -5 dB, ont été construites selon le même protocole.

2.1 Matériel et méthodes

2.1.1 Les locuteurs

Pour chaque langue manipulée, plusieurs femmes et hommes ont été enregistrés dans leur langue native pour générer les masqueurs de parole. La production des items cibles a été réalisée par une femme de langue maternelle française différente de celles enregistrées pour les masqueurs français.

2.1.2 Les participants

Vingt huit volontaires ont participé à l'Expérience 1 et 30 à l'Expérience 2. Aucun sujet n'a participé aux deux expériences. Tous étaient des étudiants de langue maternelle française, âgés de 18 à 30 ans. Ils n'étaient pas familiers des langues étrangères manipulées et ne présentaient ni troubles auditifs ni troubles du langage. Ils ont été dédommagés pour leur participation.

2.1.3 Masqueurs paroliers

Afin de constituer les masqueurs de parole dans chaque langue manipulée, deux voix de femme et deux voix d'homme ont été sélectionnées parmi celles enregistrées. Chaque voix a d'abord été enregistrée individuellement dans une salle insonorisée. Tous les locuteurs ont lu dans leur langue native les mêmes textes traduits par des professionnels. De ces enregistrements ont été extraites des séquences de 4 sec selon le protocole suivant : (i) dans chacune des séquences les silences devaient être inférieurs à 500 ms, afin d'éviter que les mots cibles n'apparaissent dans un « silence » (notons cependant que cette hypothèse est peu probable puisqu'ensuite quatre séquences sont mixées), (ii) suppression des phrases avec une prononciation incorrecte, une prosodie exagérée. Les signaux ont été normalisés à 70 dB-A puis mixés par quatre afin d'obtenir les masqueurs de parole.

2.1.4 Masqueurs de bruit fluctuant

Contrairement aux masqueurs paroliers, constitués d'informations linguistiques et acoustiques, les masqueurs de bruit fluctuant ne possèdent que des informations acoustiques. Comparer les performances obtenues avec ces deux types de masqueurs permettra de mettre en évidence l'effet des informations linguistiques de chaque langue manipulée. A l'aide du logiciel MATLAB, les masqueurs de bruit fluctuant ont été construits à partir des masqueurs de parole dans chaque langue manipulée (pour le détail de la procédure, voir Hoen et al., 2007).

2.1.5 Items cibles

Quatre-vingt un mots français bisyllabiques ainsi que 81 pseudo-mots ont constitué les items cibles. Ces 162 items cibles étaient identiques dans les deux expériences. Les mots

ont été sélectionnés dans la base de données Lexique2 (New et al., 2004) et ont été équilibrés en fréquence d'occurrence (de 0,29 à 175,65 par million de mots ; moyenne = 17,16 ; ET = 30,43). Les pseudo-mots ont été construits en respectant les règles phonotactiques de la langue française, comme par exemple *trouchet*.

2.1.6 Stimuli

Pour chaque expérience, les 162 stimuli correspondaient aux 81 mots et 81 pseudo-mots mixés à un masqueur (parolier ou bruit fluctuant). Les items cibles étaient toujours insérés à 2,5 sec du début d'un masqueur. Parmi les 81 mots cibles, 3 ont été utilisés pour la phase d'entraînement. Une moitié des 78 mots restants a été mixée avec des masqueurs de parole, l'autre moitié avec des masqueurs de bruit fluctuant. Le même protocole a été appliqué pour les 81 pseudo-mots, 3 ont été gardés comme items d'entraînement, 39 ont été insérés dans des masqueurs paroliers et les 39 restants avec des masqueurs de bruit fluctuant.

Au total, 6 listes expérimentales différentes ont été générées. Chacun des 156 items (78 mots et 78 pseudo-mots) n'était présent qu'une seule fois dans chacune des listes. Chaque participant ne voyait qu'une seule liste, soit chaque participant ne rencontrait les items cibles qu'une seule fois. Au final, les listes étaient composées de 13 mots et de 13 pseudo-mots présentés dans chacune des 6 conditions expérimentales (2 types de masqueurs (paroliers vs. bruit fluctuant), 3 langues (Expériences 1 et 2 : gaélique irlandais vs. italien vs. français)). A travers toutes les listes, l'ensemble des stimuli ont été présentés dans toutes les conditions. A l'intérieur de chaque liste, l'ordre de présentation des items était randomisé pour chaque passation.

Dans l'Expérience 1, les items cibles ont été insérés dans les masqueurs avec un RSB de 0 dB et dans l'Expérience 2 avec un RSB de -5 dB.

2.1.7 Procédure

Les participants étaient testés individuellement. Ils étaient assis dans une salle d'expérimentations insonorisée face à un écran d'ordinateur. Les stimuli étaient présentés avec Eprime (Psychology Software Tools, Inc., Pittsburg, PA) et délivrés de manière diotique à travers un casque audio (Beyerdynamic DT 48, 200 Ω) à 65 dB SPL (niveau d'écoute confortable). Les participants devaient réaliser une tâche de décision lexicale sur les items cibles insérés dans les masqueurs. Il leur était demandé de répondre le plus vite et le plus correctement possible si l'item cible était un mot ou pas en appuyant sur l'une des deux touches pré sélectionnées sur le clavier. Les participants ne pouvaient entendre les items cibles qu'une seule fois. Ils passaient eux-mêmes d'un essai à l'autre en appuyant sur la barre espace. Avant la phase test, une phase d'entraînement était proposée afin de s'habituer au mode de présentation des stimuli ainsi qu'à la voix cible. Il s'agissait de 12 essais, chacun des 3 mots et 3 pseudo-mots sélectionnés pour l'entraînement a été présenté dans un masqueur de parole et dans un masqueur de bruit fluctuant dans une des trois langues manipulées. La durée de l'expérience était de 30 min environ.

3 Résultats

Pour les deux expériences, les moyennes de Temps de Réaction (TR : intervalle de temps, en ms, entre la présentation de l'item cible et la pression du bouton) ainsi que les taux d'erreurs ont été mesurés. Pour l'analyse des TR, les essais pour lesquels les participants ont commis des erreurs (Expérience 1 : 22,1% ; Expérience 2 : 34,1%), ou n'ont pas répondu dans le temps imparti de 4500 ms (Expérience 1 : 0,8% ; Expérience 2 : 4,3%) n'ont pas été inclus. Les essais dont les TR étaient inférieurs à 300 ms (Expérience 2 : 0,2%) n'ont pas été pris en compte.

Deux analyses de variance à mesures répétées (ANOVA) ont été conduites sur l'Expérience 1 à 0 dB et sur l'Expérience 2 à -5 dB. La première ANOVA a considéré comme variable dépendante les TR et comme variables intra-sujets le facteur Nature des masqueurs (masqueurs paroliers vs. masqueurs de bruit fluctuant) et le facteur Langue des masqueurs (Expériences 1 et 2 : irlandais vs. italien vs. français). Dans la seconde ANOVA, les taux d'erreurs constituaient la variable dépendante, les variables intra-sujets étant les mêmes que dans la première ANOVA.

3.1 Expérience 1 : 0 dB

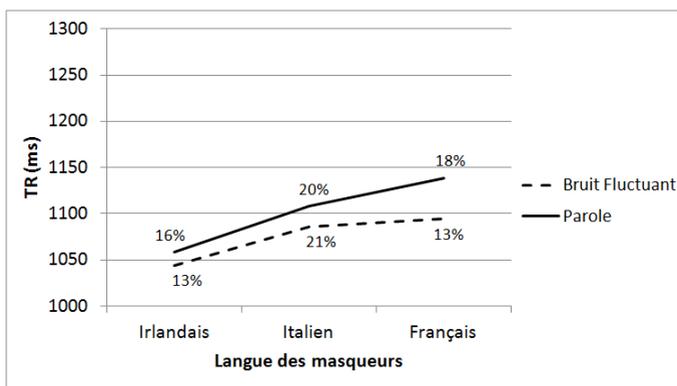


FIGURE 1 – Effet de la Nature des masqueurs (parolier vers bruit fluctuant) sur les TR en fonction de la Langue des masqueurs à 0 dB. Les valeurs indiquées sur les courbes représentent les taux d'erreurs pour chaque condition expérimentale.

Les résultats sont présentés dans la Figure 1. L'effet de la Nature des masqueurs n'est pas ressorti significatif à l'analyse de variance ($F(1,27)=3,13$; $p=.09$). En revanche, l'ANOVA a révélé un effet significatif de la Langue des masqueurs ($F(2,54)=7,12$; $p<.001$). Les participants ont été plus rapides lorsque les masqueurs étaient produits en irlandais (1051 ms), ils ont été plus lents avec les masqueurs italiens (1097 ms) et français (1117 ms). L'interaction entre ces deux facteurs n'a pas été significative

($F(2,54)=0,3$; $p=.73$). Les comparaisons post-hoc réalisées avec le test HSD de Tukey ont montré qu'avec les masqueurs de bruit fluctuant, aucune différence entre les langues n'a été révélée. Avec les masqueurs paroliers, seule une différence existe entre l'irlandais et le français, c'est - à - dire entre les langues qui ont conduit aux performances extrêmes. Enfin, pour chacune des langues, l'effet de la Nature des masqueurs n'a pas été significatif, aucune différence dans les TR n'a été observée entre les deux types de masqueurs (paroliers versus bruit fluctuant).

L'ANOVA conduite sur les taux d'erreurs des participants n'a révélé aucun effet significatif.

Dans cette expérience, la situation d'écoute à 0 dB n'a pas permis de mettre en lumière l'effet des informations linguistiques des langues manipulées puisque les deux types de masqueurs ont entraîné des performances équivalentes. Ce résultat suggère que les effets observés (TR significativement différents entre les masqueurs paroliers irlandais et français) seraient essentiellement dus aux différences acoustiques entre les langues. Nous avons donc décidé de rendre la tâche plus difficile en modifiant le RSB à -5 dB, afin de déterminer si des interférences de nature linguistique pourraient être révélées.

3.2 Expérience 2 : -5 dB

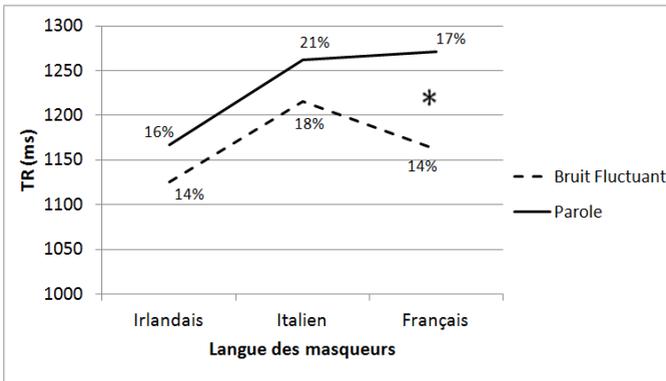


FIGURE 2 – Effet de la Nature des masqueurs (parolier vers bruit fluctuant) sur les TR en fonction de la Langue des masqueurs à -5 dB. Les valeurs indiquées sur les courbes représentent les taux d'erreurs pour chaque condition expérimentale.

Les résultats sont présentés dans la Figure 2. Contrairement à l'Expérience 1, nous avons observé un effet significatif de la Nature des masqueurs ($F(1,29)=8,92$; $p=.005$). Les TR étaient plus longs lorsque les masqueurs étaient composés de parole (1234 ms) plutôt que de bruit fluctuant (1167 ms). L'analyse a également révélé un effet significatif de la Langue des masqueurs ($F(2,58)=5,83$; $p<.005$). En moyenne, les participants ont été

plus rapides avec les masqueurs en gaélique irlandais (1146 ms), ils ont été plus lents avec de l'italien (1217 ms) ou du français (1239 ms) en bruit de fond. L'interaction entre ces deux facteurs n'a pas été significative ($F(2,58)=1,25$; $p=.29$). Des comparaisons post-hoc effectuées avec le test HSD de Tukey ont montré qu'avec les masqueurs de bruit fluctuant, aucune différence significative entre les langues n'a été observée. En revanche, avec les masqueurs paroliers, des différences significatives ont été révélées entre l'irlandais et le français, ainsi qu'entre l'irlandais et l'italien. Enfin, l'effet de la Nature des masqueurs n'a été présent que pour le français, avec des TR significativement plus rapides pour les masqueurs de bruit fluctuant que pour les masqueurs paroliers.

Concernant les taux d'erreurs des participants, aucun effet significatif n'est ressorti de l'analyse de variance.

Pour l'irlandais et l'italien, l'absence de différence entre les TR obtenus avec les masqueurs paroliers et les masqueurs de bruit fluctuant suggère que l'effet de masque de ces langues est d'origine acoustique. Pour le français, cette différence est significative et met en lumière le rôle des informations linguistiques de cette langue lorsqu'elle est présentée en fond sonore.

4 Discussion

Dans cette étude, nous nous sommes intéressés à la compréhension de la parole dans la parole et plus précisément aux interférences impliquées dans cette situation. Le but était de comparer l'effet de masque de langues différentes (italien et irlandais) de la langue de la parole cible par rapport à l'effet de masque d'une langue identique (français).

Dans la situation de parole dans la parole, nous avons observé que la parole cible a été masquée différemment selon la langue des masqueurs. Ce résultat n'a été observé qu'à -5 dB. A 0 dB, la ségrégation des flux entre les masqueurs paroliers et la parole cible étant plus facile, les interférences entre ces deux sources sonores seraient réduites.

A -5 dB, l'identification des items cibles français a été significativement plus facile avec de l'irlandais qu'avec de l'italien ou du français dans les masqueurs paroliers. L'italien a conduit à un niveau de dégradation équivalent à celui du français (langue identique à la parole cible). Néanmoins, l'utilisation du bruit fluctuant nous a permis d'observer que l'effet de masque de ces deux langues ne serait pas de même nature. En effet, pour l'italien, les performances n'ont pas été significativement différentes entre les masqueurs de bruit fluctuant (composés uniquement d'informations acoustiques) et les masqueurs paroliers (composés d'informations acoustiques et linguistiques). Cette absence de différence montre que les informations linguistiques de l'italien n'ont pas ou peu participé à la dégradation de la parole cible contrairement aux informations acoustiques qui ont eu un rôle prépondérant. Avec du français, la dégradation de la parole cible a été significativement plus importante avec les masqueurs paroliers qu'avec le bruit fluctuant. Ainsi, le rôle des informations linguistiques du français a été mis en lumière. Ces résultats sont en accord avec ceux observés précédemment. A l'aide d'une tâche de retranscription de mots cibles français dégradés par des masqueurs paroliers en français à quatre, six ou huit locuteurs, Hoen et al. (2007) ont montré qu'avec quatre locuteurs en bruit de fond, des informations lexicales seraient encore identifiables. En effet, parmi les erreurs des participants se trouvaient des mots présents dans les masqueurs paroliers.

L'irlandais a eu comme l'italien un effet de masque de nature acoustique. Aucune différence de performances entre les deux types de masqueurs (paroliers versus bruit fluctuant) n'a été observée. Mais ces langues, toutes les deux différentes de la langue de la parole cible, n'ont pas affecté la parole cible avec autant d'efficacité. L'irlandais a eu un effet de masque plus faible que celui de l'italien. Des analyses acoustiques sur les masqueurs italiens et irlandais seraient nécessaires afin de déterminer quelles sont les informations acoustiques qui jouent un rôle prépondérant dans le masquage de la parole cible.

5 Conclusion

Nos résultats ont clairement montré que les interférences observées dans la situation de la parole dans la parole étaient purement acoustiques pour les langues différentes de la parole cible (italien et irlandais) et à la fois acoustiques et linguistiques lorsque la parole cible et la parole concurrente étaient produites dans la même langue. Il serait intéressant de répliquer ce travail en manipulant d'autres langues différentes de la langue de la parole cible afin d'observer la nature de leur effet de masque (purement acoustique ou linguistique et acoustique).

Remerciements

Cette recherche a été financée par l'European Research Council (projet SpiN N°209234) et par un financement de la Délégation Générale de l'Armement.

Références

- BRONKHORST, A. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica*, 86: 117-128.
- BRUNGART, D.S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109(3): 1101-1109.
- HOEN, M., MEUNIER, F., GRATALOU, C., PELLEGRINO, F., GRIMAULT, N., PERRIN, F., PERROT, X., and COLLET, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech communication*, 49: 905-916.
- MADDISON, I., FLAVIER, S., MARSICO E. and PELLEGRINO, F. (2011). LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. <http://www.lapsyd.ddl.ish-lyon.cnrs.fr/>. [consulté le 04/04/2012].
- NEW, B., PALLIER, C., BRYBAERT M. and FERRAND, L. (2004). A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3): 516-524.
- RHEBERGEN, K.S., VERSFELF, N.J. and DRESCHLER, W.A. (2005). Release from informational masking by time reversal of native and non naive interfering speech. *Journal of the Acoustical society of America*, 118(3): 1274-1277.
- VAN ENGEN, K.J. and BRADLOW, A.R. (2007). Sentence recognition in native- and foreign language multi-talker background noise. *Journal of the Acoustical Society of America*, 121(1): 519-526.