Influence des paramètres psycholinguistiques du cocktail party sur la compréhension d'un signal de parole cible

Claire Grataloup[1], Michel Hoen[1, 2], François Pellegrino[1] & Fanny Meunier[1]

1 Laboratoire Dynamique Du Langage CNRS UMR 5596-Université Lyon2
Institut de Sciences de l'Homme 14 avenue Berthelot 69363 LYON Cedex 07

2 Laboratoire Neurosciences et Systèmes Sensoriels
CNRS UMR 5020, Université Claude Bernard, Lyon, France

claire.grataloup@univ-lyon2.fr

ABSTRACT

This paper presents results from an experiment studying the cognitive ability to understand a speech signal in a babble background noise. We further tested subject's sensitivity to characteristics of target and competitor words. Our results show that words are better reconstructed than pseudowords. Intelligibility of words is not influenced by a change (number of voices, frequency of words) in the background babble noise whereas intelligibility of pseudowords is. Pseudowords perception is easier when words that constitute the background noise are low frequency words and when the number of voices is fewer.

1. INTRODUCTION

La perception du langage parlé est une tâche complexe, menée quotidiennement, et qui implique des fonctions cognitives de haut degré. La plupart du temps, l'écoute du signal de parole est endommagée par des artefacts qui perturbent sa compréhension : la présence de bruit ambiant en est l'exemple le plus courant. Le système cognitif doit donc contourner cette difficulté en reconstruisant les portions de signal mal ou non-perçues. Plusieurs études ont montré que dans de telles circonstances la parole reste, dans une certaine mesure, intelligible [1-2-3]. Il existe donc une capacité cognitive de restauration de la parole dégradée.

Une situation particulièrement complexe à traiter et pourtant fréquente se présente lorsque le signal que nous devons percevoir est « camouflé » à l'intérieur d'un flot de paroles provenant de différents locuteurs. Bien que le message cible soit très dégradé, notre système cognitif reste capable de le restaurer de façon suffisante à ce que nous le comprenions. Ce phénomène, décrit par Cherry en 1981 [4] comme l'effet « cocktail party » a été étudié à plusieurs reprises (Bronkhorst, 2000 pour une revue [5]) pour tenter d'identifier les processus cognitifs qui permettent d'isoler la voix qui nous intéresse parmi un brouhaha sonore composé de plusieurs autres voix. Les résultats montrent que la compréhension du message cible dépend à la fois du masquage informationnel et du masquage énergétique imposés par les voix concurrentes [6]. Le masquage énergétique correspond à un recouvrement spectrotemporel même partiel du son cible et du son concurrent. Le masquage informationnel est dû à un recouvrement des informations colportées par les deux signaux.

Une étude de Hoen, Grataloup, Grimault, Perrin, Perrot, Pellegrino, Meunier et Collet (2006) [7] a récemment étudié la sensibilité des locuteurs aux caractéristiques du mélange de parole concurrent. Des mots isolés étaient présentés dans des cocktails de voix composé de 4, 6 ou 8 voix présentées à l'endroit ou inversées (reversed speech). Les résultats révèlent une meilleure performance globale pour la condition à 6 voix que pour celles à 4 et 8 voix. De plus, pour un cocktail composé de 4 voix les mots cibles sont mieux perçus lorsque la parole est inversée que lorsqu'elle est à l'endroit. Cet effet disparaît pour les cocktails 6 et 8 voix. Le masque énergétique augmentant avec le nombre de voix compris dans le cocktail, les auteurs interprètent ces résultats comme révélant qu'à 4 voix dans la condition à l'endroit le masquage informationnel est en place, les mots du cocktail pouvant être activés -ce qui n'est pas le cas pour la condition inversée- et qu'à 6 et 8 voix il disparaît, ne laissant la place qu'au masque énergétique.

Afin d'approfondir cette hypothèse d'activation lexicale des mots du cocktail, nous avons réalisé une expérience ou étaient testés : pour les cibles, le facteur type d'item (mot/pseudomot) avec pour les mots leur fréquence et leur nombre de voisins phonologiques ; et pour le bruit, le nombre de voix et la fréquence des mots qui le constituent. Cette étude mesurait la reconstruction cognitive de signaux de parole (mots et pseudomots) détériorés par la présence de voix concurrentes (cocktail).

2. EXPÉRIENCE

Le principe de l'expérience est de faire entendre à des sujets normo entendants des signaux de parole cibles présentés à l'intérieur d'un cocktail de voix concurrentes. La tâche consiste à identifier l'item cible prononcé par une voix différente de celles composant le bruit de fond.

2.1. Méthode

Matériel : Items cibles

Nous avons sélectionné à l'aide de la base *Lexique* [8] 120 noms communs de la langue française, monosyllabiques et de vocabulaire courant. Deux critères étaient contrastés : leur fréquence d'occurrence dans la langue (facteur f) et leur nombre de voisins phonologiques (facteur v). Ces deux facteurs ont été croisés de façon à construire quatre catégories de 30 mots cibles chacune : table 1. Par exemple, le mot *mage* a une faible fréquence d'occurrence mais possède beaucoup de voisins phonologiques (exemples : *cage, gage, nage, page, rage, sage*…).

Table 1 : Moyennes et fourchettes des fréquences et des voisins phonologiques utilisées

	fréquence	voisins
-	M=2.54 [1, 4.94]	M=11.68 [3, 17]
+	M=66.39 [50.1, 149.23]	M=26.35 [21, 35]

120 pseudomots monosyllabiques ont également été construits en recombinant les phonèmes des mots cibles. Les 240 items ont été enregistrés (22 kH, mono, 16 bits) par une locutrice de langue maternelle française dans un caisson insonorisé. Le matériel utilisé pour l'enregistrement se composait : du logiciel Wavelab lite version 2.53 Steinberg editor, d'une carte son digigram VX pocket440, d'un préamplificateur Behringer ultragain MIC 2000 et d'un micro Rode NT1 équipé d'une membrane Popkiller K&M. Les enregistrements ont ensuite été normalisés à -3 dB à l'aide du logiciel Adobe® Audition® 1.0.

Matériel : Cocktails

Nous avons créé 6 types de cocktails à partir de 16 enregistrements (de 12 min en moyenne) réalisés par 8 locuteurs différents (4 hommes et 4 femmes) qui lisaient chacun une liste de mots fréquents et une liste de mots peu fréquents (1250 mots par liste). Dans les deux listes contrastées en fréquence nous avons équilibré le nombre de lettres des mots, le nombre de syllabes (voir table2) et la proportion de mots de 1, 2, 3 et 4 syllabes.

Table 2 : Critères d'équilibration des listes de mots composant le fond sonore. M = moyenne, ET = écart-type.

Critère	Liste F+	Liste F-
Fréquence	M=151.25 ET=451	M= 0.45 ET=0.3
Nb lettres	M=7.45 ET=2	M=7.81 ET=1
Nb syllabes	M=2.45 ET=1	M=2.63 ET=1

Nous avons ainsi constitué des cocktails C_4 composés de 2 voix féminines et de 2 voix masculines, des cocktails C_6 composés de 3 voix féminines et de 3 voix masculines et des cocktails C_8 composés de 4 voix féminines et de 4 voix masculines. Chaque cocktail existe en 2 versions, l'une fréquente (F+) et l'autre peu fréquente (F-). Dans chaque cocktail, nous avons découpé 120 extraits d'une durée de 4 secondes chacun à l'aide du logiciel Matlab qui nous a permis également de générer les stimuli finaux en superposant chaque item cible avec un extrait de chaque cocktail.

Listes expérimentales

Chaque item cible a été superposé à un extrait de chacun des 6 cocktails existants. Au total, nous avons donc généré 240*6=1440 stimuli. Nous avons ensuite créé 6 listes de mots et 6 listes de pseudomots comportant chacune 120 items. Les 6 versions de chaque item ont été réparties dans les 6 listes et contrebalancées de façon à ce que chaque item n'apparaisse qu'une seule fois par liste.

Procédure expérimentale

Les participants étaient placés face à un écran d'ordinateur de type PC, ils portaient un casque audio (Beyerdynamic DT 48) qui diffusait les stimuli un à un en mode binaural. Une consigne spécifique soit aux mots soit aux pseudomots leur était donnée oralement puis réapparaissait à l'écran en début d'expérience. Chaque sujet a été confronté à l'une des 6 listes de mots et à l'une des 6 listes de pseudomots dont l'enchaînement était spécifique à chaque sujet. La phase de test était précédée par une phase d'entraînement.

Chaque stimulus se compose de 4 secondes de cocktail à l'intérieur duquel l'item cible apparaît 2.5 secondes après le début du bruit. Après chaque stimulus ils devaient retranscrire au clavier l'item cible perçu. La moitié des sujets a commencé par les mots et l'autre moitié par les pseudomots. Une pause était effectuée entre les deux moitiés de l'expérience qui durait 45 minutes.

Sujets

Quarante sujets (25 femmes, 15 hommes) de langue maternelle française ont passé l'expérience1. Leur âge variait entre 18 et 25 ans (moyenne=21.5 ans). Tous ont passé une audiométrie tonale confirmant une audition normale (seuils<20dB) sur la gamme de fréquence des sons de la parole humaine. Aucun d'entre eux ne souffrait de troubles du langage et tous avaient une vue normale ou corrigée. Les participants étaient naïfs quand au but de l'étude et ont été rémunérés 7.5 € chacun.

2.2. Résultats

Nous avons effectué une analyse statistique ANOVA sur les 40 sujets et 240 items en considérant comme variable aléatoire d'une part, les sujets ($F1$) et d'autre part les items ($F2$). La variable dépendante était le % de restitution entière et correcte des items par les sujets.

Effet des items cibles

On observe d'une manière très forte que les mots sont mieux restitués 61% (ET=5.56) que les pseudomots

39.42% (ET=7.02) : figure1. $F1(1,39)=294.87$; $p<.0001$; $F2(1,238)=34,73$; $p<.0001$.

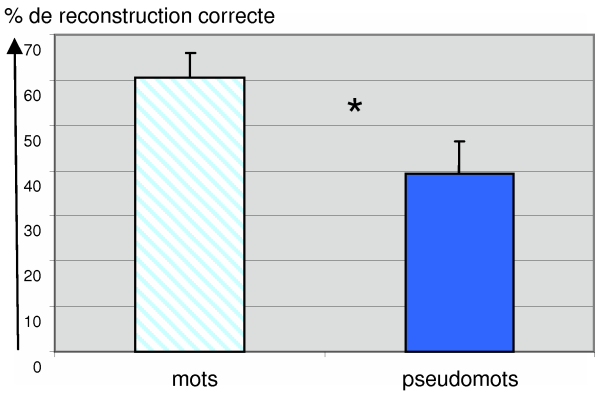

Figure 1 : Effet du type d'item cible sur le % de reconstruction. L'étoile signifie que la différence est statistiquement différente au seuil .05.

Pour les mots on observe un effet simple de leur fréquence d'occurrence. Les mots de haute fréquence sont mieux reconstruits 71% (ET=6.39) que les mots de basse fréquence 50% (ET=6.73). $F1(1,39)=368.72$; $p<.001$; $F2(1,118)=14.92$; $p<.001$. On observe également un effet simple du nombre de voisins phonologiques des mots cibles. Contrairement à ce à quoi on se serait attendu, les mots qui ont le plus de voisins sont ceux qui sont le mieux reconstruits 70% (ET=6.30) pour la condition v+ contre 51% (ET=6.95) pour la condition v- : figure2. $F1(1,39)=249.15$; $p<.0001$; $F2(1,118)=14.46$; $p<.001$.

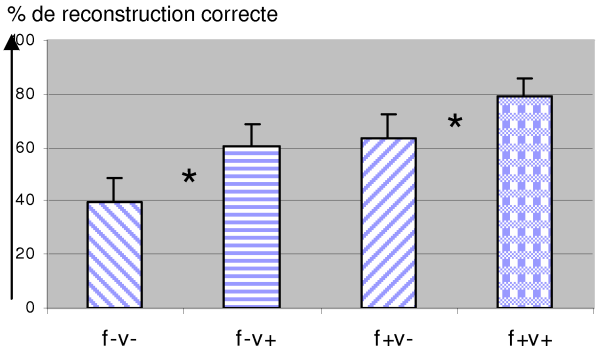

Figure 2 : Effet de la fréquence et du nombre de voisins phonologiques des mots cibles sur leur % de reconstruction.

Effet des cocktails

On n'observe pas d'effet significatif du nombre de voix des cocktails. En revanche on observe un effet de la fréquence des mots du cocktail sur la restitution des pseudomots uniquement : figure 3. En moyenne, les pseudomots sont reconstruits à 38% dans un cocktail fréquent (ET=7.85) et à 41% dans un cocktail non fréquent (ET=9.21). $F1(1,39)=3.98$; $p=.05$, $F2(1,119)=4.75$; $p<.05$. Les mots sont respectivement reconstruits à 61% (ET=7.44) et 60%(ET=7.37). $F1(1,39)=1.4$; n.s. ; $F2(1,119)=2.7$;n.s..

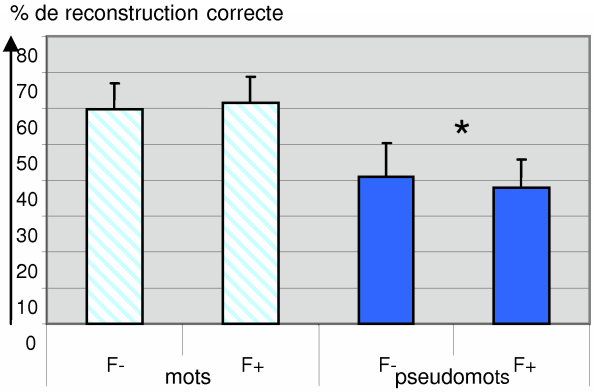

Figure 3 : Effet de la fréquence des mots du cocktail sur le % de reconstruction des items cibles.

Interaction nombre de voix et fréquence des mots du cocktail

Il faut noter que nous observons une interaction significative entre le nombre de voix et la fréquence des mots du cocktail pour la reconstruction des items. Pour les mots, cette interaction n'est significative que par items : $F1(2,78)=2.1$;n.s. ; $F2(2,238)=3.37$, $p<.05$. Le traitement des pseudomots en revanche présente une interaction significative par sujets $F1(2,78)=3.7$, $p<.05$ et par items $F2(2,238)=4.64$; $p=.01$. En moyenne, on n'observe pas de différence entre cocktails pour la condition 6 voix (42% (ET=14,1) contre 38% (ET=10.43)) ni pour la condition 8 voix (36%, ET=12.8 contre 38%, ET=12.07). Cependant, on observe une différence significative entre les % de reconstruction pour les deux types de cocktails dans la condition à 4 voix. Les pseudomots sont reconstruits à 44% (ET=13.6) dans le cocktail peu fréquent et seulement à 38% (ET=13.77) dans le cocktail fréquent : figure 4.

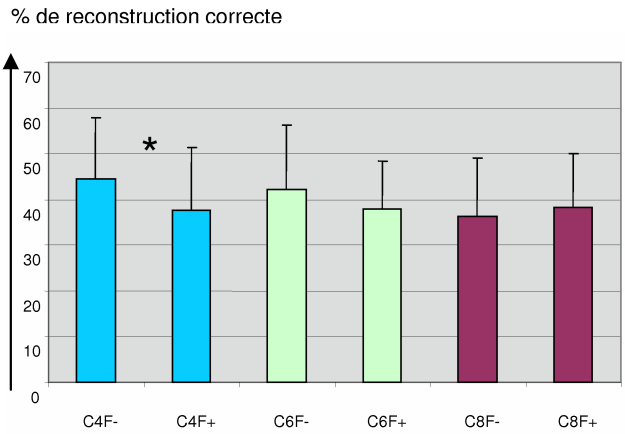

Figure 4 : Effet d'interaction entre la fréquence et le nombre de voix du cocktail sur le % de reconstruction des pseudomots cibles.

3. DISCUSSION

Les résultats montrent que les mots sont mieux reconstruits que les pseudomots ce qui est cohérent car les pseudomots n'ont pas de représentation lexicale stockée en mémoire. De ce fait, ils ne bénéficient d'aucune aide lexicale au moment de l'effort de reconstruction contrairement aux mots qui, eux, bénéficient de cette aide.

En ce qui concerne les mots, on observe un fort effet de leur fréquence d'occurrence. Les mots les plus fréquents sont mieux restitués que les mots de basse fréquence. La fréquence du mot influence les performances de restitution quelque soit le nombre de voix qui composent le bruit de fond et quelque soit la fréquence des mots du bruit de fond. L'effet de fréquence est un effet très robuste qui apparaît dans la plupart des tâches cognitives proposées aux sujets [9]. La fréquence étant une caractéristique de stockage du mot dans le lexique mental, plus sa fréquence est élevée, plus l'accès au mot est facile. On observe ici, que cet effet se retrouve lorsque la perception des mots cibles est perturbée par celle de mots concurrents.

Le résultat le plus intéressant observé dans cette expérience est sans aucun doute l'effet de fréquence du cocktail sur la restitution des mots cibles. Lorsque les mots distracteurs du cocktail sont de basse fréquence, le % de reconstruction est plus élevé. A l'inverse, si les mots du cocktail sont de forte fréquence, ils gênent la reconstruction du mot cible. Ce résultat peut-être interprété de deux façons : soit par un effet attentionnel différent selon le niveau de fréquence des mots, soit par une différence d'activation des mots des deux catégories. En d'autres termes : lorsque les mots du cocktail sont de forte fréquence, ils attirent plus l'attention du locuteur (effet de familiarité) et de ce fait les ressources attentionnelles disponibles pour traiter l'item cible sont moindres (diminution du % de reconstruction) ou bien, lorsque les mots du cocktail sont de basse fréquence, ils sont moins saillants et le système dispose donc de plus de ressources attentionnelles pour traiter le stimulus cible. L'autre explication, peut-être plus plausible, est que les mots de basse fréquence du cocktail sont moins activés et donc moins en compétition avec les mots cibles. D'autres expériences sont nécessaires afin de clarifier ce point. Cependant, il est certain que les locuteurs sont sensibles aux caractéristiques lexicales des mots du cocktail.

De plus, bien que l'on n'observe pas d'effet simple du nombre de voies du cocktail, on observe cependant une interaction entre la fréquence des mots et le nombre de voix composant le cocktail. L'effet de fréquence des mots du cocktail n'apparaît que pour la condition 4 voix : Les stimuli sont significativement mieux reconstruits dans un cocktail peu fréquent à 4 voix que dans un cocktail fréquent à 4 voix. Ce résultat suggère que c'est bien dans la condition où seulement 4 voix sont mélangées que les locuteurs peuvent être sensibles à la qualité des mots prononcés. Au-delà de 4 voix, le bruit de fond devient trop dense pour pouvoir discerner une différence de fréquence entre les mots composant les deux types de cocktails. A 4 voix cependant, le bruit de fond n'est pas encore suffisamment chargé et il est possible que les locuteurs soient influencés par un facteur lexical des mots concurrents. Ce résultat rejoint celui de Hoen et collaborateurs [7].

4. CONCLUSION

Cette étude présente les premiers résultats mettant en évidence une sensibilité des locuteurs aux caractéristiques lexicales de voix concurrentes de type cocktail lors d'une tâche de perception de parole. Ce paradigme pourrait permettre alors l'exploration des compétitions lexicales entrant en jeu dans la compréhension de la parole d'une manière plus simultanée que les paradigmes d'amorçages actuellement utilisés.

5. NOTES DES AUTEURS ET REMERCIEMENTS

Nous remercions la région Rhône-Alpes qui a permis la réalisation de cette étude grâce au projet Emergence 2004 attribué à Fanny Meunier.

BIBLIOGRAPHIE

[1] Warren, R.M. (1970). Restoration of missing speech sounds. Science, 167, 392--393.

[2] Scott, S. K., Blank, S. C., Rosen S., and Wise, R. J. S. (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400-06.

[3] Davis, M.H. & Johnsrude, I.S. (2003). Hierarchical processing of spoken language comprehension. Journal of Neuroscience 23, 3423-3431.

[4] Cherry, E. (1953). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am., 25, 975–979.

[5] Bronkhorst, A. (2000). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica 86, 117–128.Trans. Speech and Audio Proc., 7(6):697-708, 1999.

[6] Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott K.R. (2001). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am., 110, 2527–2538.

[7] Hoen, M.; Grataloup, C.; Grimault, N.; Perrin, F.; X. Perrot; Pellegrino, F.; Meunier, F.; Collet, L. Tomber le masque de l'information : effet *cocktail party*, masquage informationnel et interactions psycholinguistiques en situation de compréhension de la parole dans la parole. JEP 2006.

[8] New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. L'Année Psychologique, 101, 447-462.

[9] Segui, J., Melher, J., Frauenfelder, U., et Morton, J. (1982). The word frequency effect and lexical access. Neuropsychologia, 20(6), 615-627