Tomber le masque de l'information: effet *cocktail party*, masque informationnel et interférences psycholinguistiques en situation de compréhension de la parole dans la parole.

Michel Hoen[1,2], Claire-Léonie Grataloup[1], Nicolas Grimault[2], Fabien Perrin[2], Xavier Perrot[2], François Pellegrino[1], Fanny Meunier[1], Lionel Collet[2]

[1]Laboratoire dynamique du Langage
UMR5096 CNRS, Université Lumière, Lyon, France

[2]Laboratoire Neurosciences et Systèmes Sensoriels
UMR5020 CNRS, Université Claude Bernard, Lyon, France

michel.hoen@phonak.ch

ABSTRACT

Up to now, the comprehension of speech in noise and more particularly in concurrent speech sounds was rarely studied in the domain of psycholinguistics. In this paper we report a study testing the differential effects of speech derived noises as multi-talker cocktail party sounds and their time-reversed pendent on the comprehension of isolated words. Results suggest that different levels of linguistic information from concurrent speech signals can compete with linguistic information in target signals, mainly depending on the spectral saturation caused by increasing the number of voices in concurrent signals. These results suggest linguistically specific participations in informational masking effects occurring in the context of speech in speech comprehension.

1. INTRODUCTION

Bien que nous soyons aptes à comprendre de la parole distillée par les casques d'une chambre anéchoique, nous sommes plus souvent confrontés à la situation où la parole nous parvient du chaos acoustique d'un grand carrefour ou de l'ambiance criarde d'une salle de réunion. Pourtant, nous restons souvent capables d'en comprendre le message. La parole reste intelligible dans des conditions acoustiques extrêmement variables et malgré la présence de quantités importantes de bruits interférents. Ce phénomène, appelé « *Effet Cocktail Party* » correspond à une faculté cognitive spécialisée nous permettant de focaliser notre attention sur un flux auditif particulier parmi différents flux concurrents. Depuis sa première description par Cherry en 1953 [1], l'effet cocktail party a donné lieu à un grand nombre d'études ayant permis de cerner certains de ses fondements cognitifs. Il a été par exemple démontré que l'effet cocktail party reposait sur la capacité du système auditif humain à réaliser une séparation spatiale des flux concurrents [2]. Cependant, lorsque les flux acoustiques ont une origine spatiale commune, ou lorsque le son est présenté de façon diotique (le même signal aux deux oreilles), le système cognitif doit alors pouvoir utiliser d'autres indices.

Dans ce contexte, différentes études ont pu mettre en évidence l'importance d'indices temporels lents ou d'indices de surface, comme des différences globales de fréquence fondamentale (F_0), d'accent, de style discursif ou encore d'intensité entre les flux à séparer [3-8]. En situation de compétition de flux acoustiques, on décrit souvent les effets de masques attribuables aux flux interférents comme pouvant agir à deux niveaux principaux : un niveau énergétique et un niveau informationnel [7-8]. L'effet de masque énergétique est dû aux propriétés spectrotemporelles des sons concurrents. Il se produit lorsque la parole est émise en présence d'un bruit à large bande spectrale qui va se superposer partiellement, en temps ou en fréquence, au signal de parole cible. L'effet de masque informationnel est quant à lui attribuable au type d'information contenue dans le bruit [4-5]. Dans ce cas, ce sont les informations que les signaux véhiculent qui vont entrer en compétition et perturber l'interprétation cognitive qui sera faite du signal cible. Le phénomène de masque informationnel est particulièrement saillant lorsqu'un signal de parole est émis en présence d'autres signaux de parole. Dans ce contexte particulier, Brungart et coll., ont étudié l'intelligibilité d'un signal de parole cible en fonction du nombre de voix concurrentes (2, 3 ou 4 locuteurs) et du Rapport du Signal au Bruit (RSB) de signaux de parole interférents [7-8]. Leurs résultats ont montré une décroissance linéaire des performances en fonction du RSB, dès que trois voix au moins étaient mises en concurrence. La condition à deux voix concurrentes étant résolue sur la base d'indices de surface permettant de discriminer les voix (modulations de F_0, timbre ou style discursif). Ces études montrent la disparition de l'impact de tels indices acoustiques de surface dans la résolution de l'effet cocktail party à partir de situations d'interférence à trois voix de genre identique (ou à partir de 4 voix concurrentes de genre mixte). Ainsi, il semble qu'au-dessous de quatre voix concurrentes, la résolution de l'effet cocktail party repose essentiellement sur l'utilisation d'une stratégie de séparation spatiale des flux ou d'indices acoustiques de surface. Bien que la caractérisation de l'effet cocktail party ait donné lieu à un très grand nombre d'études, peu d'expériences se sont intéres-

sées aux interférences fines ayant lieu au-delà de 4 voix concurrentes. En particulier, les effets de masques psycholinguistiques pouvant apparaître dans les situations de compréhension de la parole dans la parole n'ont jamais été mis en évidence. Pourtant, le concept de masque informationnel prend dans ce contexte un sens particulier puisqu'il pourrait alors être attribué aux différents types d'information linguistique contenue dans la parole, telles que l'information prosodique, phonologique ou lexicale. On pourrait alors aisément imaginer, au sein du concept de masque informationnel, l'existence de sous-types d'effets liés à l'accessibilité dans le son concurrent de ces différents niveaux d'information linguistique. Cette accessibilité devrait dépendre de la transparence spectro-temporelle du signal interférent, dès lors, la sensibilité des sujets à ces différents niveaux d'information serait hypothétiquement modulée par la saturation énergétique du signal (proportionnelle au nombre de locuteurs) ainsi que par le RSB.

La présente expérience a été conduite afin de mettre en évidence, au sein du masque informationnel, l'existence de différents types d'interférences linguistiques dans des situations de compréhension de la parole dans la parole. Nous avons pour cela conduit une expérience de compréhension de mots isolés en présence de bruits paroliers. Dans ces bruits, nous avons manipulé le nombre de locuteurs et la nature physique exacte des signaux. Nous avons comparé les effets de masques dus à des enregistrements de 'cocktail party' standards à 4, 6 ou 8 voix simultanées et ceux dus aux mêmes enregistrements mais inversés selon leur dimension temporelle. L'inversion temporelle de la parole, ou 'reversed speech', a parfois été considérée comme la manipulation la plus drastique pouvant être appliquée à un signal de parole [9]. En réalité, la parole inversée conserve les propriétés énergétiques du signal de parole source, mais 'sonne' aussi comme la parole naturelle, puisqu'elle en conserve certains traits infra-segmentaux. Les voyelles en particulier sont bien conservées et certaines consonnes présentent de bons degrés de réversibilité. Mieux encore, lorsque plusieurs flux de parole inversée sont superposés, le signal composite sonne comme un signal de cocktail party, et des phonèmes peuvent y être perçus. En revanche, l'information lexicale est totalement perdue. Afin de pouvoir différencier les effets de masque linguistiques du simple effet de masque énergétique, nous avons également employé comme son masquant un bruit large bande ayant les mêmes propriétés spectrotemporelles qu'un bruit de cocktail party mais ne contenant pas d'information linguistique.

2. MATERIELS ET METHODES

2.1. Bruits Interférents

Bruits de *cocktail party* multi-locuteurs

Trois bruits de cocktail party ont été créés en mixant 4, 6 ou 8 enregistrements de locuteurs uniques. Chaque locuteur a été enregistré individuellement à l'aide d'une chaîne d'acquisition et de numérisation comportant un microphone Røde NT1, un préamplificateur Ultragain MIC 2000 et une carte son RolandED UA-30, les sons étant numérisés à 44 kHz sur 16 bits. Chaque source individuelle consistait en une voix masculine ou féminine prononçant des phrases intelligibles en langue française, dont les noms propres avaient été supprimés. Les voix sources ont toutes subi la même chaîne de traitement: i) suppression des pauses et silences excédant une seconde; ii) suppression des portions d'enregistrements contenant des erreurs de prononciation ou des marques prosodiques inappropriées; iii) réduction du bruit de fond optimisée pour les signaux de parole (CoolEdit Pro© 1.1 – Dynamics Range Processing – preset Vocal limiter); iv) calibration en dBA et normalisation de chaque source à 80 dBA (Larson Davis System LD824 et oreille artificielle: AEC101); et enfin : v) mixage des différentes sources et sauvegarde au format .wav (44kHz, 16 bits, Stéréo).

Bruits de *cocktail party* inversés

Les bruits de cocktail party inversés ont été obtenus en appliquant une inversion point à point selon l'axe temporel des bruits de cocktails multi-locuteurs décrits ci-dessus. Ceci fut réalisé grâce à une routine implémentée dans le logiciel Adobe® Audition™ dans sa version 1.0.

Bruit large bande associé

Afin d'obtenir un bruit à large spectre aux propriétés énergétiques semblables à celles de nos bruits paroliers nous avons décidé de partir du son cocktail party comprenant 8 locuteurs (celui-ci ayant le spectre énergétique le plus large et le plus dense) et d'en dériver un bruit ne contenant plus aucune information linguistique. Pour ce faire, nous avons commencé par extraire l'enveloppe temporelle du bruit original sous 60Hz, afin d'en dériver les fluctuations dynamiques lentes. Puis, par une transformée de Fourier (FFT), nous avons calculé l'énergie spectrale du signal d'origine et en avons extrait la distribution des phases. Les phases ont été redistribuées de façon aléatoire, puis réinjectées dans l'enveloppe temporelle du bruit de cocktail party original. Enfin, l'énergie globale rms du bruit obtenu a été ajustée à celle du signal original. Le bruit résultant possède la même énergie spectrale et la même enveloppe que le bruit original, mais les phases étant aléatoires, il ne comporte plus aucune information d'ordre linguistique.

2.2. Mots Cibles

320 mots français monosyllabiques, tri-phonémiques ont été enregistrés. Ils ont été sélectionnés dans une gamme de fréquence d'occurrence moyenne (0.19 occurrences par million (opm) à 146.71 opm ; moy = 20.96 ; DS = 21.37) d'après Lexique2 [10], ceci afin d'éviter des items de trop haute ou trop basse fréquence. Les mots isolés étaient prononcés par un locuteur masculin unique et enregistrés en chambre sourde à l'aide d'un microphone Sony ECM-MS907 et sauvegardés au format .wav (44 kHz, stéréo, 16 bits).

2.3 Conditions et Sujets

L'effet de masque de 7 types de bruits sur la compréhension de mots cibles isolés a été testé: trois bruits de cocktail party à 4, 6 et 8 voix mixtes, trois bruits de cocktails inversés à 4, 6 et 8 voix mixtes et un bruit à large bande. Chacun de ces bruits était testé pour des RSBs de -3, 0, +3 et +6 dBs, générant un total de 28 conditions. 36 sujets ont pris part à l'expérimentation. Nous avons généré 36 listes de 8 mots cibles, équilibrées en fréquence et en nombre de voisins phonologiques. Les mots dans les listes finales avaient une fréquence de 3.92 opm (DS = 0.006) et un nombre de voisins phonologiques de 19.83 (DS = 0.06). Les participants étaient tous étudiants, âgés de 18 à 32 ans, de langue maternelle française et dépourvus de déficits auditifs ou langagiers diagnostiqués, ils étaient rémunérés pour leur participation.

2.4 Stimuli

Les stimuli étaient 288 fichiers au format .wav (44 kHz, 16bits, stéréo) ayant chacun une durée de 4 secondes. Le bruit de fond était présent durant toute la durée du stimulus et le mot cible était systématiquement inséré à 2.5s du début du fichier. Les mots cibles avaient des durées variables de 242.68 ms à 914.47 ms (moy = 550.32 ms, DS = 134.56 ms). Les extraits de bruits ont été sélectionnés au hasard et contrebalancés. Par ailleurs, comme le mixage final du bruit et des mots cibles résultait en des variations d'intensité globale des stimuli, nous avons appliqué suite au mixage, une normalisation en intensité aléatoire sur une gamme de $^{+}/_{-}$ 3 dB par pas de 1dB. L'intensité globale des stimuli obtenus ne pouvait ainsi être prédictive de la condition de stimulation.

2.5 Procédure

Les participants étaient assis face à un écran d'ordinateur. Les stimuli étaient présentés de façon diotique au moyen d'un casque audio (Beyerdynamic DT 48, 200Ω) à un niveau d'écoute confortable fixé individuellement. Tous les bruits, listes de mots et conditions étaient aléatoirisées parmi les sujets. La tâche consistait à écouter les stimuli et à retranscrire au clavier le mot cible ou bien la portion de mot cible entendu. Les sujets étaient familiarisés au décours temporel des essais sur 12 exemples. La durée totale de l'expérience allait de 30 à 60 min en fonction de l'habileté des sujets à utiliser un clavier informatique. Les transcriptions des sujets étaient enfin analysées en termes de proportion de mots correctement reproduits.

3. RESULTATS

Afin de tester l'effet du nombre de voix présentes dans un bruit de cocktail party interférent sur la compréhension d'un signal de parole cible, nous avons réalisé une première Analyse de variance (Anova) en prenant les taux de récupération lexicale individuels comme variable dépendante.

L'analyse incluait les facteurs de bruit: 'Type' (2: Cocktail Party ou Cocktail Inversé), 'Nombre' de voix présentes dans le bruit (3: 4, 6 ou 8) et le RSB (4 : -3, 0, +3 et +6dB), (voir Fig. 1). Cette analyse a révélé un effet principal uniquement pour le facteur RSB ($F(3,105) = 185,31$; $p<.05$), le taux de récupération lexicale décroissant de façon linéaire avec le RSB dans notre gamme de RSBs testés. Les deux autres facteurs, 'Type' et 'Nombre' n'avaient pas d'effets principaux significatifs, respectivement $F(1,35) = 1,18$; n.s. et $F(2,70) = 2,19$; n.s., suggérant une absence d'effet global de ces facteurs. En revanche, l'interaction de second ordre entre ces deux facteurs était significative ($F(2,70) = 4,53$; $p<.05$). Aucune autre interaction n'était significative, en particulier, le facteur RSB n'interagissait ni avec le facteur 'Type' ($F(3,105) = 0,93$; n.s.), ni avec le facteur Nombre ($F(6,210) = 0,50$; n.s.), suggérant une décroissance globale des performances avec le RSB indépendante de la condition de bruit considérée. Le facteur RSB a pour cette raison été ensuite supprimé des autres analyses en moyennant l'ensemble des données obtenues aux mêmes RSBs. Un test post-hoc de type LSD ($\alpha=.05$), a ensuite été appliqué à l'interaction de second ordre significative afin d'établir l'influence du nombre de locuteurs sur la façon dont les différents types de bruits masquaient les mots cibles. Au sein des bruits paroliers de type cocktail party, cette analyse révélait un effet non linéaire du nombre de locuteurs puisque le bruit à 6 voix avait l'effet de masque le moins important et différait significativement des effets de masques des bruits de cocktail à 4 voix ($p=0.004$), et à 8 voix ($p=0.056$). Les deux bruits de cocktails à 4 et 8 voix avaient les effets de masque les plus importants. En comparant les effets de masques dus aux bruits de type cocktail party à ceux dus aux bruits de type cocktail inversé, une différence significative était observée entre le cocktail à 4 voix et le cocktail inversé à 4 voix ($p=0.004$), le bruit de cocktail à 4 voix étant associé à l'effet de masque le plus important. Bien que l'effet de masque dû aux bruits de cocktails inversés semble décroître avec le nombre de voix impliquées, cette décroissance est trop lente pour aboutir à des différences significatives entre bruits de cocktail inversés à 4, 6 ou 8 voix, nous considérerons donc que ces trois bruits ont des effets de masques très comparables.

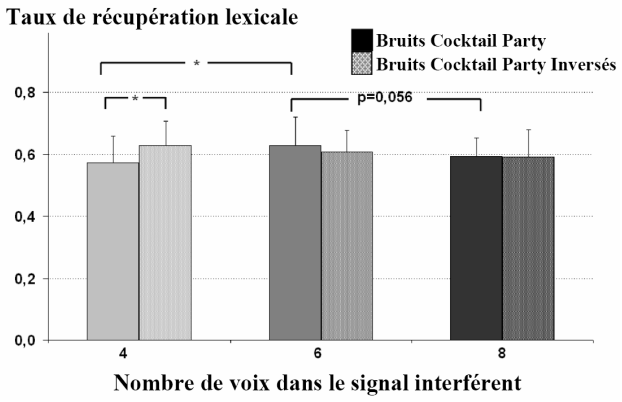

Figure 1: *Taux de récupération de mots isolés en fonction du type de bruit parolier interférent.*

Dans l'analyse précédente nous avons observé une différence significative entre les effets de masques dus au bruit cocktail party à 4 voix et au même bruit inversé. Cette observation suggèrerait l'existence de deux niveaux de masquage informationnel dans une condition à 4 locuteurs, uniquement si ces deux effets sont eux-mêmes distincts d'un effet de masque purement énergétique.

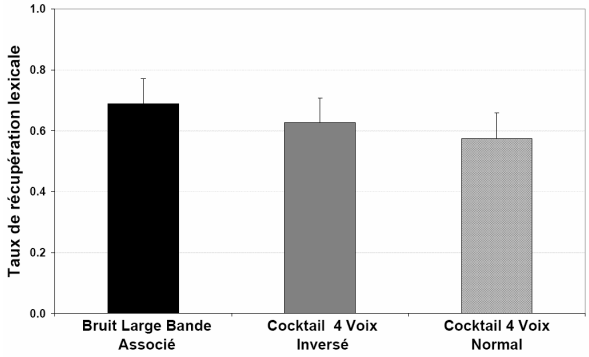

Figure 2: *Taux de récupération de mots isolés pour trois types de bruits concurrents: le bruit associé, le cocktail party inversé à 4 voix et le cocktail party à 4 voix.* ***(Toutes différences significatives)***

Afin d'élucider cette question nous avons réalisé une seconde Anova à un niveau, en prenant la moyenne des taux de récupération lexicale obtenus au même RSB comme variable dépendante et pour seul facteur le 'Type' de bruit (3: Bruit associé, Cocktail Party 4 voix inversé et Cocktail Party 4 voix). Cette analyse révélait un effet principal significatif du 'Type' de bruit considéré ($F(2,70)=23,45$; $p<.05$) et les comparaisons planifiées étaient toutes significatives, montrant que ces trois types de bruits avaient des effets de masque significativement différents. Le bruit large bande associé avait le moins grand effet de masque, puis venait le bruit de cocktail party à 4 voix inversé et enfin le bruit de cocktail party à quatre voix (voir Fig. 2).

4. DISCUSSION

Les résultats de cette expérience sur l'effet de masque informationnel en situation de cocktail party multi-locuteur ont permis de mettre en évidence, au sein de l'effet de masquage informationnel, de potentiels effets de l'accessibilité de différentes sources d'information linguistique dans un bruit parolier concurrent. En effet, nous avons pu montrer qu'un bruit physique associé, un bruit de cocktail inversé et un bruit de cocktail party standard ayant tous trois des propriétés énergétiques comparables avaient trois effets de masquage distincts, le signal de cocktail party normal ayant l'effet de masquage le plus fort. Ces trois niveaux de masque pourraient être attribués respectivement à un niveau énergétique pur (bruit large bande associé), à l'ajout d'un niveau d'ordre phonologique (bruit de cocktail inversé) et enfin à un effet combiné énergétique, phonologique et lexical dans le cas du bruit de cocktail party à 4 voix standard. L'accessibilité d'information lexicale dans le cocktail à 4 voix étant attestée par certaines erreurs des sujets ayant répondu, rarement mais de façon symptomatique, avec des mots issus du bruit de cocktail plutôt qu'avec les mots cibles. Cet effet de masque lexical disparaît dans notre expérience pour des nombres de locuteurs supérieurs à 4 voix, le cocktail party à 6 locuteurs étant même le bruit de cocktail le moins masquant. Ceci pourrait être dû à une disparition de l'accessibilité de l'information lexicale dans le bruit de cocktail party du fait d'une progressive saturation spectrale du signal causée par l'ajout progressif de locuteurs. Ce résultat inédit offre un nouveau cadre d'étude pour les mécanismes de compétition d'informations ayant lieu lors de l'accès au lexique mental puisqu'il met pour la première fois en évidence des compétitions d'informations psycholinguistiques dues au contenu informationnel d'un signal interférent. Ce paradigme pourrait permettre de tester directement des hypothèses sur les phénomènes de compétition d'information lexicale dans le cadre de l'accès au lexique telles que décrites dans les modèles psycholinguistiques.

REMERCIEMENTS

Cette étude a été réalisée grâce aux fonds de l'ACI (n° 67068) du *Ministère de l'Enseignement Supérieur et de la Recherche français,* attribuée à Fanny Meunier.

BIBLIOGRAPHIE

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

[2] Bronkhorst, A. (2000). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' *Acustica*. 86, 117–128.

[3] Egan, J., Carterette, E., and Thwing, E. (1954). ''Factors affecting multi channel listening,'' *J. Acoust. Soc. Am.* 26, 774–782.

[4] Dirks, D., and Bower, D. (1969). ''Masking effects of speech competing messages,'' *J. Speech Hear. Res.* 12, 229–245.

[5] Festen, J., and Plomp, R. (1990). ''Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing,'' *J. Acoust. Soc. Am.* 88, 1725–1736.

[6] Darwin, C., and Hukin, R. (2000). ''Effectiveness of spatial cues, prosody and talker characteristics in selective attention, *J. Acoust. Soc. Am.* 107, 970–977.

[7] Brungart, D. (2001a). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' *J. Acoust. Soc. Am.* 109, 1101– 1109.

[8] Brungart, D. (2001b). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' *J. Acoust. Soc. Am.* 110, 2527– 2538.

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

[10] New, B., Pallier, C., Brysbaert, M., Ferrand, L. Lexique 2: A New French Lexical Database (In Press) *Behavior Research Methods, Instruments, & Computers.*