

10ème Congrès Français d'Acoustique

Lyon, 12–16 Avril 2010

Masquage auditif temps-fréquence avec des stimuli de forme Gaussienne

Thibaud Necciari, Sophie Savel, Sabine Meunier, Richard Kronland-Martinet, Sølvi Ystad

Laboratoire de Mécanique et d'Acoustique (CNRS–UPR 7051), 31, chemin Joseph Aiguier – 13402 Marseille Cedex 20

necciari@lma.cnrs-mrs.fr

Etre capable d'adapter la représentation temps-fréquence (T-F) d'un son à la perception auditive de ce dernier est un enjeu de grande importance pour de nombreuses applications (*e.g.*, réalité virtuelle, synthèse sonore, codage audio). Afin de répondre à une telle problématique, le masquage auditif – qui se définit comme l'élévation du seuil d'audibilité d'une source sonore (la *cible*) en présence d'une autre source (le *masque*) – a été mesuré dans le plan T-F à l'aide de stimuli de forme Gaussienne. De tels signaux minimisent l'incertitude T-F et n'activent qu'une fenêtre d'observation spectro-temporelle du système auditif (van Schijndel *et coll.*, 1999, *J. Acoust. Soc. Am.* 105, 3425–3435). Masque et cible étaient des sinusoïdes modulées par une fenêtre Gaussienne brève (9,6 ms) et à support fréquentiel compact ("Equivalent Rectangular Bandwidth" = 600 Hz). Le masque avait une fréquence porteuse fixée à 4 kHz et un niveau fixé à 60 dB SL. Les résultats obtenus lorsque seule la distance temporelle ou fréquentielle masque-cible est manipulée sont semblables à ceux obtenus dans la littérature avec des stimuli de localisation T-F comparable. En revanche, les résultats obtenus pour différentes combinaisons T-F sont nouveaux et constituent une base cruciale pour l'implémentation des données de masquage dans un outil de traitement du signal permettant d'éliminer l'information inutile – d'un point de vue perceptif – dans les représentations T-F des sons. En supposant une combinaison linéaire des effets de masquage dans les domaines temporel et fréquentiel, deux modèles de prédiction sont testés. Les résultats témoignent de l'importance d'une approche T-F des phénomènes de masquage et indiquent que ceux-ci ne se limitent pas à la simple combinaison des effets dans les domaines temporel et fréquentiel.

1 Introduction

Le masquage auditif se définit comme l'élévation du seuil d'audibilité d'une source sonore (la "cible") en présence d'une autre source (le "masque"). En quantifiant le degré d'élévation du seuil de détection de la cible en présence du masque on mesure la "quantité de masquage". Dans la littérature, le masquage a été étudié pour des conditions dans lesquelles masque et cible sont présentés simultanément (masquage fréquentiel), ou non-simultanément (masquage temporel) [1]. En revanche, peu d'études traitent du masquage dans le plan temps-fréquence (T-F) [2, 3]. En masquage fréquentiel, la séparation fréquentielle entre masque et cible (ΔF) est variée. Dans la méthode courante des patterns de masquage, la fréquence du masque (F_M) est fixe et la quantité de masquage est mesurée pour différentes fréquences de cibles (F_C). Ces études emploient généralement des stimuli de longue durée (≥ 300 ms) afin de conserver une bande passante étroite. En masquage temporel, F_M et F_C sont identiques et la séparation temporelle masque-cible (ΔT) est variée. Le masquage *antérieur* (la cible précède le masque) se différencie du masquage *postérieur* (le masque précède la cible) [4]. Afin de contrôler les propriétés temporelles des signaux, ces études emploient généralement des stimuli large bande.

L'absence de données de masquage T-F semble provenir de l'impossibilité de contrôler les signaux de façon conjointe en temps et en fréquence, en raison de l'incertitude T-F [5]. Un signal de forme Gaussienne minimise cette incertitude et fournit donc la meilleure localisation dans le plan T-F. Par conséquent, le choix de tels signaux pour la mesure du masquage T-F semble approprié. De plus, à la différence des signaux longs et/ou large bande couramment employés dans les études de masquage, des signaux Gaussiens brefs (≤ 10 ms) sont

supposés n'activer qu'un nombre limité de fenêtres d'observation T-F du système auditif [6].

Dans cette étude, le masquage T-F a été mesuré à l'aide de sinusoïdes modulées par une fenêtre Gaussienne de courte durée et à support fréquentiel compact. Masque et cible étaient identiques. Dans un premier temps, seule la séparation fréquentielle (Exp. 1) ou temporelle (Exp. 2) masque-cible a été variée. Les résultats obtenus sont comparés à ceux de la littérature pour des signaux de localisations T-F variées. Dans l'Expérience 3, le masquage a été mesuré pour différentes combinaisons $\Delta F/\Delta T$. D'un point de vue théorique, ces résultats fournissent des données nouvelles de masquage T-F pour des signaux courts et à bande étroite. D'un point de vue pratique, ces résultats constituent une base cruciale pour le développement d'un outil de traitement de signal permettant de prédire les effets de masquage dans les sons complexes.

2 Méthode générale

2.1 Stimuli

Masque et cible étaient des sinusoïdes modulées par une fenêtre Gaussienne [6]

$$s(t) = \sqrt{\Gamma} \sin\left(2\pi f_0 t + \frac{\pi}{4}\right) e^{-\pi(\Gamma t)^2} \quad (1)$$

où f_0 est la fréquence porteuse et Γ permet de contrôler les caractéristiques spectro-temporelles du signal $s(t)$. Γ a été fixé à 600 de manière à obtenir des signaux de bande passante ("Equivalent Rectangular Bandwidth", $ERB = \Gamma = 600$ Hz) et de durée ("Equivalent Rectangular Duration", $ERD = \Gamma^{-1} = 1,7$ ms) constantes. Le déphasage de $\pi/4$ permet d'obtenir des signaux de même

énergie quelle que soit f_0 . La fenêtre Gaussienne étant à support infini, les signaux ont été tronqués dans le domaine temporel à l'aide d'une fenêtre de Tukey de 9,6 ms. Cette fenêtre permet de conserver les propriétés de la Gaussienne. Sa durée a été déterminée par une procédure d'optimisation numérique permettant d'éliminer les lobes spectraux secondaires [7]. Les niveaux SPLs des Gaussiennes étaient calculés à partir de la tension RMS des signaux.

2.2 Procédure

Tous les seuils ont été mesurés par une procédure adaptative à choix forcé à 3 intervalles avec un critère "3-down 1-up" permettant d'estimer le point à 79,4% de réponses correctes sur la fonction psychométrique. Un essai consistait à présenter au sujet 3 intervalles de 200 ms indiqués visuellement sur le boîtier de réponse. Le masque était présenté dans les trois intervalles. La cible était présentée aléatoirement avec le masque dans un des trois intervalles. Le sujet devait alors indiquer dans quel intervalle il pensait avoir entendu la cible. À la fin de chaque essai, la bonne réponse était notifiée visuellement au sujet. Le niveau initial de la cible était supérieur de 10–15 dB au seuil estimé. Au cours d'une série, le niveau de la cible variait de façon adaptative par pas de 5 dB jusqu'à la seconde inversion, puis 2 dB au-delà. Une série se terminait après 12 inversions et le seuil obtenu correspondait à la moyenne arithmétique des niveaux de la cible aux 10 dernières inversions. Si l'écart-type de ces 10 inversions était supérieur à 5 dB, le seuil était éliminé. Deux estimations de seuils étaient obtenues pour chaque condition expérimentale. Si l'écart-type de ces deux mesures était supérieur à 3 dB, jusqu'à quatre mesures supplémentaires étaient obtenues. Le seuil final correspondait à la moyenne de toutes les estimations. Dans toutes les expériences, les sujets pratiquaient de 1 à 2 sessions de 30 min par jour. Six séries (soit 6 estimations de seuil) étaient pratiquées par session.

2.3 Sujets

Six sujets normo-entendants ont participé aux Expériences 1 et 2. Quatre des 6 sujets (S1–S4) ont participé à l'Expérience 3. Tous avaient des seuils inférieurs à 15 dB HL dans la bande de fréquences 125 Hz–8 kHz.

2.4 Dispositif expérimental

Tous les stimuli étaient générés en temps réel par un processeur numérique (Tucker-Davis Technologies, TDT System III) doté d'une résolution de 24 bits et cadencé à 48 kHz. Le processeur (TDT RP2.1) était piloté par un programme DELPHI à partir d'un PC. Masque et cible étaient calculés dans DELPHI, générés par le processeur puis envoyés sur deux canaux distincts en sortie du processeur. Lorsqu'un bruit de fond continu était requis pour masquer les produits de combinaison, un bruit large bande était généré et filtré passe-bas (-96 dB/octave) par le processeur, puis envoyé sur le même canal que le masque. Les deux canaux de sortie étaient acheminés vers deux convertisseurs numérique/analogique. Les signaux analogiques étaient alors atténués (TDT PA5) et sommés (TDT SM3) avant d'être envoyés à l'amplificateur (TDT HB7) puis à l'écouteur droit d'un casque (Sennheiser HD545) étalonné au tympan. Chaque sujet était testé individuellement dans une cabine audiométrique.

3 Expérience 1 : masquage fréquentiel

3.1 Conditions expérimentales

La fréquence du masque (F_M) était fixée à 4 kHz et son niveau (L_M) à 60 dB SL. Masque et cible étaient présentés simultanément ($\Delta T = 0$). Les patterns de masquage ont été mesurés pour 11 valeurs de ΔF définies sur l'échelle ERB [8] : 0, ± 1 , ± 2 , ± 3 , ± 4 , ± 5 et ± 6 ERBs ($F_C = 2521, 2833, 3181, 3568, 4000, 4480, 5015, 5611, 6274, 7012$ et 7835 Hz).

Pour éviter que la détection des produits de combinaisons ne produise des irrégularités dans les patterns de masquage, un bruit de fond continu a été ajouté lorsque $F_C > F_M$ [9]. Le niveau et la fréquence de coupure du bruit étaient déterminés de façon à masquer le produit de combinaison le plus proéminent : le "Cubic Difference Tone" (CDT) dont la fréquence (F_{CDT}) est $2F_M - F_C$. La fréquence de coupure du bruit a été choisie comme la borne supérieure de l'ERB centré sur F_{CDT} . Le niveau du bruit a été fixé à 50 dB SPL. Pour vérifier que la présence du bruit de fond n'affectait pas la détection des signaux, la condition $\Delta F = 0$ a été mesurée avec et sans bruit sur 5 sujets (S1–S5). Dans cette condition, aucun produit de combinaison n'est généré donc les seuils mesurés avec et sans bruit devraient être identiques. Ce fût effectivement le cas pour tous les sujets (différences de seuils < 3 dB).

Chaque session contenait des conditions mesurées avec ($F_C \geq F_M$) ou sans ($F_C < F_M$) bruit de fond. Lors d'une session, F_C était choisie aléatoirement. Afin de déterminer la quantité de masquage à chaque ΔF , les seuils de détection dans le silence pour les 11 cibles Gaussiennes ont été mesurés pour chaque sujet dans une expérience préliminaire. Les seuils obtenus à 4 kHz ont également permis de fixer les niveaux SL individuels des masques (72–75 dB SPL).

3.2 Résultats et discussion

La Figure 1 présente les patterns de masquage individuels et moyen. Premièrement, on remarque que seul le sujet S2 obtient un pic à $\Delta F = 0$. Les autres sujets obtiennent un creux (S1, S3 et S4) ou un plateau (S5 et S6). Ceci est probablement dû à la spécificité de la condition " ΔT et $\Delta F = 0$ " dans laquelle masque et cible, des signaux identiques, sont présentés simultanément. Ainsi, le seul indice sur lequel les sujets peuvent se baser pour détecter la cible est l'incrément d'intensité perçu dans l'intervalle "masque plus cible". En notant I_M l'intensité du masque, I_{M+C} l'intensité de "masque plus cible" et en calculant la différence d'intensité relative $\Delta I/I = (I_{M+C} - I_M)/I_M$ à partir des seuils mesurés dans l'Expérience 1, on obtient une valeur moyenne de -1,6 dB. Ceci est comparable au seuil de discrimination d'intensité de -2 dB obtenu avec des sinusoïdes de 2 ms [10]. Cette condition reflète donc une tâche de discrimination d'intensité et non une condition de masquage. La présence d'un pic pour S2 peut être due au fait que ce sujet a le seuil de discrimination d'intensité le plus élevé ($\Delta I/I = +1$ dB).

Deuxièmement, pour tous les sujets et $|\Delta F| \geq 2$ ERBs, la quantité de masquage diminue avec l'augmentation de $|\Delta F|$ et on remarque que cette décroissance est plus abrupte pour $F_C < F_M$ que pour $F_C > F_M$. Pour vérifier cette asymétrie, une régression linéaire a été calculée pour chaque côté des patterns de masquage en excluant le point $\Delta F = 0$ (droites dans la

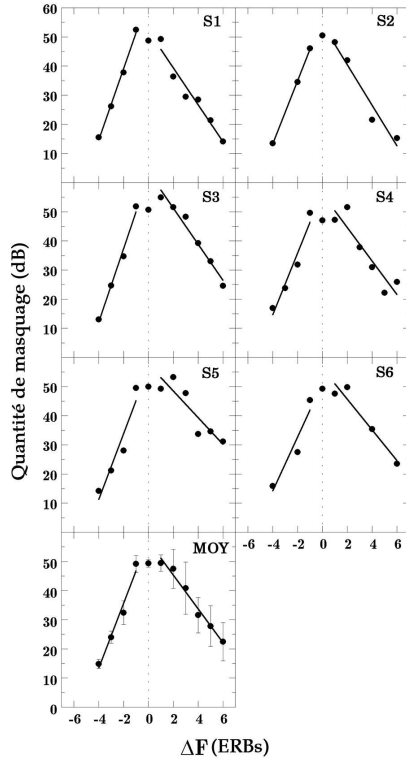


FIGURE 1 – Patterns de masquage individuels avec une régression linéaire calculée pour chaque côté (point à $\Delta F = 0$ exclu, voir texte). Le panneau inférieur indique les données moyennes avec ± 1 écart-type.

Fig. 1). Seul S2 obtient un pattern symétrique. Les données moyennes révèlent des pentes de +60 dB/octave à gauche et -39 dB/octave à droite ($r^2 \geq 0,91$). Cette asymétrie est conforme à celle rapportée précédemment dans les études de masquage fréquentiel [11, 12]. Elle s'explique par l'augmentation de la bande passante des filtres auditifs avec leur fréquence centrale [1] et par les effets de suppression qui dominent pour $F_C > F_M$ et pour $L_M > 50$ dB SPL [13]. Le fait que S2 ait obtenu un pattern symétrique peut être dû au fait que ce sujet nécessite un plus haut niveau de masque pour produire l'asymétrie [11, 13]. Une comparaison des résultats de l'Expérience 1 avec des données de la littérature obtenues avec des masques de fréquences et de niveaux comparables au masque Gaussien mais ayant des caractéristiques T-F différentes a révélé que l'étendue spectrale du pattern de masquage moyen de la Figure 1 est (1) plus étroite que celle produite par des masques large bande (*e.g.*, bruits [14], sinusoïdes tronquées dans le domaine temporel avec une fenêtre rectangulaire produisant un large étalement spectral [15]), (2) plus large que pour des masques sinusoïdaux stationnaires pour lesquels l'énergie est concentrée autour de F_M [11] et (3) comparable à celle produite par des masques sinusoïdaux tronqués avec des fenêtres permettant de limiter l'étalement spectral [12].

4 Expérience 2 : masquage temporel

4.1 Conditions expérimentales

Masque et cible avaient la même fréquence ($F_C = F_M = 4$ kHz). Le niveau du masque était fixé à 60 dB SL. La séparation temporelle entre masque et cible (ΔT), définie "pic à pic", était 0, 5, 10, 20 ou 30 ms. Le mas-

quage antérieur n'a pas été mesuré car nous avons supposé ses effets négligeables étant donné le niveau (72–75 dB SPL) et la brièveté de notre masque [4, 16, 17]. Ceci a été confirmé par une expérience pilote. Au cours d'une session, la valeur de ΔT était choisie aléatoirement.

4.2 Résultats et discussion

Les résultats sont présentés dans la Figure 2. En moyenne, le masquage décroît de 50 dB pour $\Delta T = 0$ à 6 dB pour $\Delta T = 30$ ms. Cette décroissance très rapide du masquage est sûrement due à la brièveté du masque Gaussien [16, 17]. L'échelle logarithmique permet de faire apparaître une décroissance linéaire du masquage rapportée dans la plupart des études antérieures [4, 16, 17, 18]. En effet, les données pour $\Delta T > 0$ peuvent être décrites par l'équation

$$AM = a \log(\Delta T) + b \quad (2)$$

où AM est la quantité de masquage, a la pente de la décroissance temporelle du masquage et b l'ordonnée à l'origine ($\log(\Delta T) = 0$). Un ajustement de l'Equation (2) aux données de la Figure 2 par la méthode des moindres carrés pondérés (le poids de chaque point était donné par l'inverse de la variance de la mesure) a fourni des valeurs de pente de -36 à -14 dB/ $\log(\Delta T)$ ($r^2 \geq 0,95$). La variabilité dans les valeurs de pente est due à la grande variabilité inter-sujets observée à $\Delta T = 5$ ms. En excluant le point à 5 ms lors du calcul de a la variabilité dans les valeurs de pente a été divisée par deux. Cet aspect est discuté ci-dessous.

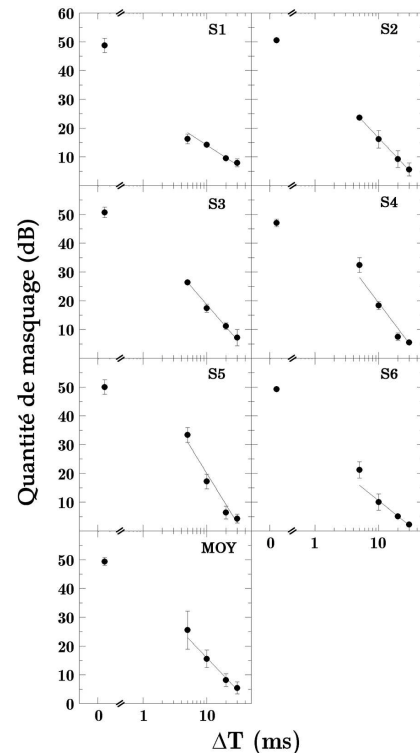


FIGURE 2 – Quantité de masquage en fonction de ΔT sur une échelle logarithmique. Les données pour $\Delta T > 0$ ont été ajustées par une droite (voir Eq. (2)) selon un critère de moindres carrés pondérés. Les barres d'erreur indiquent ± 1 écart-type.

Nos résultats sont compatibles avec une interprétation du masquage postérieur basée sur la décroissance temporelle de l'excitation produite par le masque sur la

membrane basilaire (MB) et dans le nerf auditif, ainsi que par la persistance de l'activité neurale générée par le masque dans le système central [4, 18, 1]. Pour des ΔT très petits induisant un recouvrement temporel des excitations dues au masque et à la cible sur la MB, le masquage est attribué aux mécanismes du masquage simultané. Ce recouvrement peut être accentué par la résonance des filtres auditifs [16]. Ce phénomène peut expliquer nos résultats obtenus aux courts ΔT et notamment la grande variabilité observée à 5 ms. Comme ΔT était défini "pic à pic", masque et cible se recouvraient partiellement dans cette condition. La variabilité peut donc être attribuée aux différences individuelles de résonance des filtres auditifs (seuils plus élevés = résonance plus longue) [16], ou aux différences de résolution temporelle (seuils plus bas = meilleure résolution) [19].

5 Expérience 3 : masquage temps-fréquence

5.1 Conditions expérimentales

Dans cette expérience, le masquage a été mesuré pour 32 combinaisons de $\Delta T \times \Delta F$. La fréquence et le niveau du masque étaient inchangés ($F_M = 4$ kHz, $L_M = 60$ dB SL). Bien que l'effet des produits de combinaison soit généralement ignoré en masquage postérieur, un bruit de fond continu identique à celui de l'Expérience 1 a été utilisé lorsque $F_C > F_M$, ceci en raison des petites valeurs de ΔT employées. Les seuils pour $\Delta F = 0$ ont été mesurés avec et sans bruit de fond pour tous les sujets et ΔT . Les différences de seuils se sont révélées < 3 dB, indiquant que le bruit n'affecte pas la détection des signaux. Deux groupes de conditions ont été constitués : mesures avec ($\Delta F = 0, +1, +2, +4$ et $+6$ ERBs) et sans bruit ($\Delta F = 0, -1, -2$ et -4 ERBs). Ensuite, des blocs ont été formés contenant les conditions ΔT affectées à chaque ΔF . L'ordre de passage des groupes et des blocs était aléatoire. Au cours d'une session, ΔF était fixe et ΔT était choisi aléatoirement.

5.2 Résultats

La Figure 3 présente les résultats en fonction de ΔF avec ΔT en paramètre. Pour tous les sujets et ΔF , le masquage est maximal dans la condition simultanée. Ce masquage décroît rapidement avec l'augmentation de ΔT . Notamment, pour $\Delta T > 10$ ms le masquage est inférieur à 10 dB. Pour vérifier si les patterns de la Figure 3 deviennent plus larges avec l'augmentation de ΔT , les facteurs de qualité à -3 dB, Q_{3dB} , ont été estimés pour les patterns à $\Delta T = 0, 5$ et 10 ms [2]. Les valeurs moyennes respectives sont $Q_{3dB} = 12, 3$ et 2. Celles-ci indiquent une nette diminution de la sélectivité et donc un élargissement des patterns avec le ΔT .

On remarque également que le creux observé à $\Delta F = 0$ pour $\Delta T = 0$ disparaît pour $\Delta T > 0$. En particulier, S1 et S3 révèlent un pic à +1 ERB au lieu de 0. D'autre part, les patterns demeurent asymétriques pour tous les ΔT . Des mesures répétées d'analyse de variance effectuées sur la quantité de masquage pour les facteurs côté ($F_C < F_M$; $F_C > F_M$), ΔF ($\pm 1, \pm 2$ et ± 4 ERBs) et ΔT (0, 5 et 10 ms) ont révélé un effet significatif du côté ($p < 0,001$), une interaction significative entre côté et ΔT ($p < 0,02$), mais pas d'interaction entre côté et ΔF ($p = 0,18$). Ceci suggère que la décroissance temporelle du masquage pour $F_C < F_M$ est plus rapide que pour $F_C > F_M$. Les tests statistiques ont également révélé un effet significatif de ΔF ($p < 0,001$), de ΔT ($p < 0,001$),

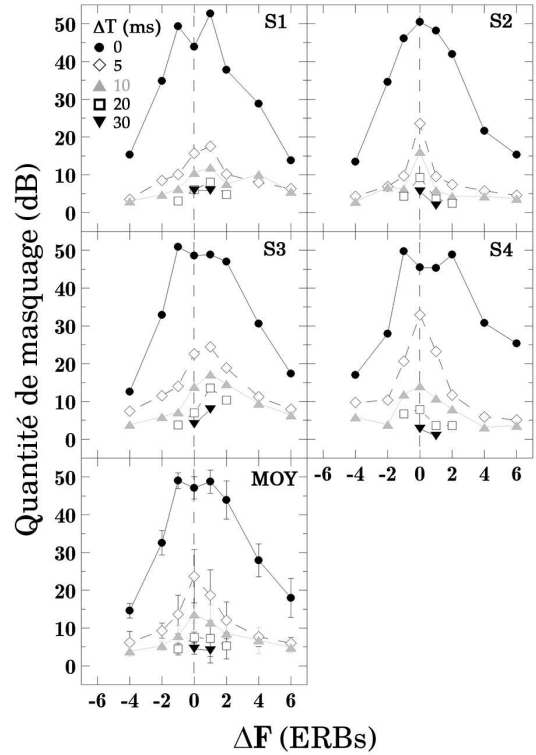


FIGURE 3 – Patterns de masquages obtenus pour 5 valeurs de ΔT .

ainsi qu'une interaction significative entre ΔF et ΔT ($p < 0,001$). Ceci suggère une activité complexe des phénomènes de masquage dans le plan T-F.

Dans la Figure 4, les résultats sont tracés en fonction de ΔT avec ΔF en paramètre. Pour tous les sujets et ΔT , le masquage diminue avec l'augmentation de $|\Delta F|$. Comme dans la Figure 2, l'échelle logarithmique permet de mettre en évidence la décroissance linéaire du masquage. En effet, pour $\Delta F = 0, \pm 1$ et ± 2 ERBs (≥ 3 points de mesure), les données pour $\Delta T > 0$ ont été ajustées par l'Equation (2) selon un critère de moindres carrés pondérés ($r^2 \geq 0,94$). Pour tous les sujets, la pente diminue avec l'augmentation de $|\Delta F|$ (en moyenne, la pente varie de -21 à -14 dB/log(ΔT) pour $F_C \leq F_M$, et de -21 à -11 dB/log(ΔT) pour $F_C \geq F_M$). On montre également que les droites obtenues pour différents ΔF convergent vers un même point sur l'axe ΔT . En calculant pour chaque sujet et ΔF le point d'intersection moyen, donné par $10^{-b/a}$, on obtient 58,2 ms pour S1 (écart-type = 14,6 ms), 46,7 ms pour S2 (6,7 ms), 63 ms pour S3 (31,4 ms) et 3,8 ms pour S4 (4 ms). Les faibles écart-types obtenus pour S1, S2 et S4 suggèrent que la décroissance temporelle du masquage atteint sa valeur limite à une valeur de ΔT qui est indépendante de ΔF .

5.3 Discussion et conclusions

A la connaissance des auteurs, seules deux études ont mesuré le masquage pour différentes combinaisons de ΔF et ΔT [2, 3]. Malgré l'emploi de masques sinusoïdaux stationnaires ($F_M = 1$ kHz, durées = 250–300 ms), ces études ont rapporté des résultats comparables à ceux de la Figure 3. Notamment (1) une décroissance rapide du masquage fréquentiel avec l'augmentation de ΔT , (2) des patterns plus larges pour $\Delta T > 0$ que pour $\Delta T = 0$ et (3) des patterns asymétriques pour tous les ΔT . Ces deux études ont également indiqué qu'une variation du ΔF n'affecte pas la caractéristique linéaire de la décroissance du masquage postérieur en fonction de $\log(\Delta T)$,

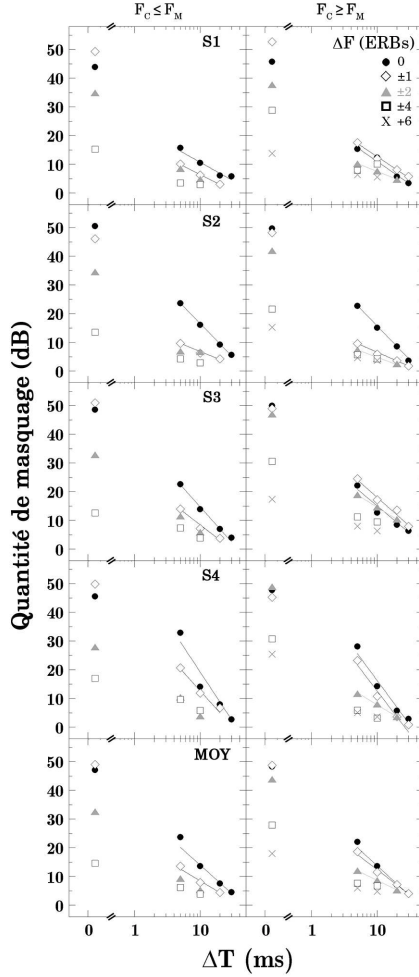


FIGURE 4 – Quantité de masquage en fonction de ΔT sur une échelle logarithmique avec ΔF en paramètre. Pour $\Delta F = 0, \pm 1$ et ± 2 ERBs, les données pour $\Delta T > 0$ ont été ajustées par une droite (voir Eq. (2)).

que la pente de cette décroissance diminue avec l’augmentation de $|\Delta F|$, et que pour un sujet donné le masquage atteint 0 dB pour une valeur de ΔT indépendante de ΔF . Ceci est en parfait accord avec les résultats de la Figure 4.

Les patterns de masquage mesurés en conditions non-simultanées sont supposés refléter la décroissance temporelle de l’excitation du masque produite sur la MB et dans le nerf auditif [2, 3]. D’après cette hypothèse, l’élargissement des patterns avec l’augmentation du ΔT est liée au fait que la raideur (ou “vitesse de décroissance”) du masquage postérieur dépend du ΔF . L’observation d’un pic à $+1$ ERB a également été rapportée dans des études antérieures [2] et pourrait être due à une asymétrie prononcée des patterns chez certains sujets. Cependant, nos résultats ne permettent pas de conclure quant à l’origine de cet effet. Mentionné dans la Section 4.2, le recouvrement temporel des excitations dues au masque et à la cible sur la MB peut être mis en cause dans les résultats de l’Expérience 3 pour de petites valeurs de ΔT et ΔF [16, 3].

Afin de vérifier si des données de masquage T-F (Exp. 3) peuvent être déduites de données mesurées indépendamment dans les domaines fréquentiel (Exp. 1) et temporel (Exp. 2), deux modèles de prédiction ont été testés. Ces derniers supposent une combinaison linéaire des effets de masquage dans les deux domaines et sont basés sur l’idée que l’étalement T-F du masquage généré par un masque peut être décrit par l’exci-

tation initiale produite par ce masque sur la MB suivie de la décroissance temporelle de cette excitation. En notant $AM(\Delta T, \Delta F)$ la quantité de masquage produite par un masque sur une cible dans le plan T-F ($\Delta T > 0, \Delta F \neq 0$), le Modèle A prédit simplement que

$$AM(\Delta T, \Delta F) = AM(0, \Delta F) - (AM(0, 0) - AM(\Delta T, 0)) \quad (3)$$

où $AM(0, \Delta F)$ représente l’excitation initiale (Fig. 1) à laquelle est soustraite la décroissance temporelle du masquage pendant la durée ΔT (Fig. 2). Les patterns de masquage individuels et moyens prédits avec le Modèle A pour $\Delta T = 5, 10$ et 20 ms sont tracés dans la Figure 5. Cette dernière révèle que pour tous les ΔT le Modèle A surestime le masquage pour $|\Delta F| \leq 2$ ERBs et sous-estime le masquage aux ΔF plus grands. L’inefficacité du Modèle A est sûrement due au fait qu’il ne prend pas en compte la décroissance linéaire du masquage postérieur en fonction de $\log(\Delta T)$, ni la dépendance au ΔF de cette décroissance. Un second modèle est donc proposé (Modèle B), basé sur l’analogie qui peut être faite entre les courbes de masquage postérieur obtenues pour différents ΔF (Fig. 4) et les résultats de masquage postérieur pour différents niveaux [17, 18]. Ceci laisse supposer que la quantité de masquage pour toute combinaison de $\Delta F/\Delta T$ peut être prédite à partir des résultats mesurés pour $\Delta F = 0$ en considérant l’action d’un masque postérieur de niveau plus faible. Ce modèle devient valable si les deux hypothèses suivantes sont vérifiées : (1) la quantité de masquage pour $\Delta T = 0$ et $\Delta F = 0$ varie de façon linéaire avec le niveau et (2) le décours temporel du masquage est une fonction linéaire de $\log(\Delta T)$ qui atteint 0 dB de masquage en un même point pour tous les niveaux. Les résultats de la Figure 4 vérifient l’hypothèse (2). L’hypothèse (1) peut être admise si on considère que la condition $\Delta T = 0$ et $\Delta F = 0$ correspond à un seuil de discrimination d’intensité (voir Sec. 3.2) qui varie de façon linéaire avec le niveau. Le Modèle B est donc défini par

$$AM(\Delta T, \Delta F) = AM(0, \Delta F) - a' \log(\Delta T) \quad (4)$$

où $a' = AM(0, \Delta F)/\log(\Delta T_{0dB})$, ΔT_{0dB} étant l’intersection avec l’axe ΔT vers laquelle convergent les fonctions de masquage postérieur. ΔT_{0dB} est déduit des valeurs de a et b estimées par l’ajustement de l’Equation (2) aux données de la Figure 2. $AM(0, \Delta F)$ est déterminé de manière identique au Modèle A. Les patterns de masquage prédits par le Modèle B pour $\Delta T = 5, 10$ et 20 ms sont présentés dans la Figure 5. Bien que la forme des patterns prédits par le Modèle B soit plus proche des données que celle issue du Modèle A, le Modèle B surestime le masquage dans toutes les conditions, en particulier pour $|\Delta F| \leq 2$ ERBs. Les erreurs du Modèle B peuvent être dues à une estimation incorrecte des fonctions de masquage postérieur en fonction du niveau [18], ou au fait que le seuil de discrimination d’intensité pour des signaux Gaussiens brefs ne varie pas de façon linéaire avec le niveau [20]. D’autre part, pour $|\Delta F| > 2$ ERBs et $\Delta T > 0$, le masquage est généralement inférieur à 10 dB. En raison des effets “plancher”, données et modèles doivent être comparés avec précaution dans ces conditions.

Finalement, les deux modèles testés ne se sont pas révélés efficaces dans la prédiction de données de masquage T-F pour des signaux courts et à bande étroite. Ceci indique qu’une combinaison linéaire des effets de

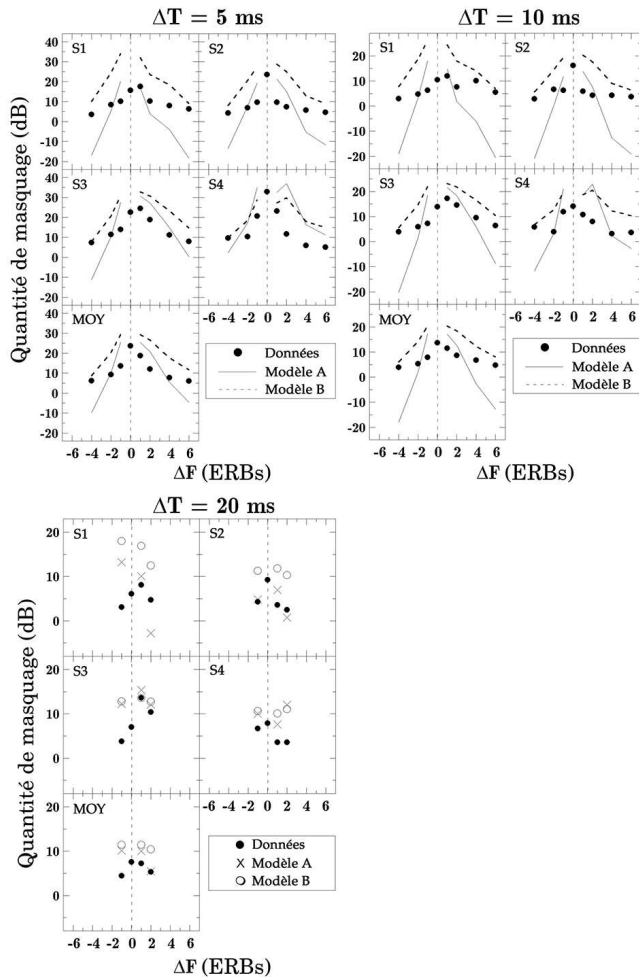


FIGURE 5 – Les patterns de masquage individuels et moyens prédits avec les Modèles A et B pour $\Delta T = 5, 10$ et 20 ms sont comparés aux données mesurées ($\bullet$).

masquage temporel et fréquentiel ne permet pas de prédire des données de masquage T-F. Des modèles de prédiction basés sur des effets d'interaction plus complexes entre masquage temps et fréquence pourraient peut-être fournir de meilleurs résultats. Il faut également tenir compte du fait que les deux approches testées ici sont simples et ne prennent pas en compte les non-linéarités du système auditif.

Ces données constituent une base cruciale pour le développement d'un outil de traitement de signal permettant de prédire les effets de masquage dans les sons complexes. En effet, chaque son réel peut être décomposé en une matrice T-F formée de composantes Gaussiennes avec des amplitudes et phases appropriées [5]. La connaissance des effets de masquage entre les composantes individuelles de cette matrice permettrait ainsi d'identifier les composantes *utiles* d'un point de vue perceptif. La mise en œuvre d'un tel outil est un enjeu de grande importance pour de nombreuses applications telles que le codage audio ou les outils d'analyse et de synthèse sonores.

Remerciements

Nous tenons à remercier Michel Jevaud pour son aide lors de la programmation des expériences. Ces travaux ont été en partie financés par Egide (PAI "Amadeus" WTZ 1/2006) et l'ANR (projet "SenSons").

Références

- [1] Moore B.C.J. *An introduction to the psychology of hearing*. Fifth ed. Academic Press, London (2003).
- [2] Kidd Jr. G., Feth L.L. "Patterns of residual masking". *Hear. Res.* 5 (1), (1981) 49–67.
- [3] Soderquist D.R., Carstens A.A., Frank G.J.H. "Backward, simultaneous, and forward masking as a function of signal delay and frequency". *J. Aud. Res.* 21, (1981) 227–245.
- [4] Elliott L.L. "Backward and forward masking of probe tones of different frequencies". *J. Acoust. Soc. Am.* 34 (8), (1962) 1116–1117.
- [5] Gröchening K. *Foundations of time-frequency analysis*. First ed. Birkhäuser, Boston (2001).
- [6] van Schijndel N.H., Houtgast T., Festen J.M. "Intensity discrimination of Gaussian-windowed tones : Indications for the shape of the auditory frequency-time window". *J. Acoust. Soc. Am.* 105 (6), (1999) 3425–3435.
- [7] Depalle P., Hélie T. "Extraction of spectral peak parameters using a short-time Fourier transform modeling and no sidelobe windows". *Proceedings of the IEEE WASPAA '97 meeting* (1997). Mohonk, NY, USA.
- [8] Moore B.C.J., Glasberg B.R. "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns". *J. Acoust. Soc. Am.* 74 (3), (1983) 750–753.
- [9] Goldstein J.L. "Auditory nonlinearity". *J. Acoust. Soc. Am.* 41 (3), (1967) 676–689.
- [10] Florentine M. "Level discrimination of tones as a function of duration". *J. Acoust. Soc. Am.* 79 (3), (1986) 792–798.
- [11] Ehmer R.H. "Masking patterns of tones". *J. Acoust. Soc. Am.* 31 (8), (1958) 1115–1120.
- [12] Bacon S.P., Viemeister N.F. "Simultaneous masking by gated and continuous sinusoidal maskers". *J. Acoust. Soc. Am.* 78 (4), (1985) 1220–1230.
- [13] Vogten L.L.M. "Simultaneous pure-tone masking : The dependence of masking asymmetries on intensity". *J. Acoust. Soc. Am.* 63 (5), (1978) 1509–1519.
- [14] Bilger R.C., Hirsh I.J. "Masking of tones by bands of noise". *J. Acoust. Soc. Am.* 28 (4), (1956) 623–630.
- [15] Green D.M. "Masking with continuous and pulsed sinusoids". *J. Acoust. Soc. Am.* 46 (4), (1969) 939–946.
- [16] Duifhuis H. "Consequences of peripheral frequency selectivity for nonsimultaneous masking". *J. Acoust. Soc. Am.* 54 (6), (1973) 1471–1488.
- [17] Penner M.J. "Effect of masker duration and masker level on forward and backward masking". *J. Acoust. Soc. Am.* 56 (1), (1974) 179–182.
- [18] Jesteadt W., Bacon S.P., Lehman J.R. "Forward masking as a function of frequency, masker level, and signal delay". *J. Acoust. Soc. Am.* 71 (4), (1982) 950–962.
- [19] Eddins D.A., Green D.M. "Temporal integration and temporal resolution". *Hearing*, second ed., 207–242. Academic Press, San Diego (1995).
- [20] Nizami L., Reimer J.F., Jesteadt W. "The intensity-difference limen for Gaussian-enveloped stimuli as a function of level : Tones and broadband noise". *J. Acoust. Soc. Am.* 110 (5), (2001) 2505–2515.