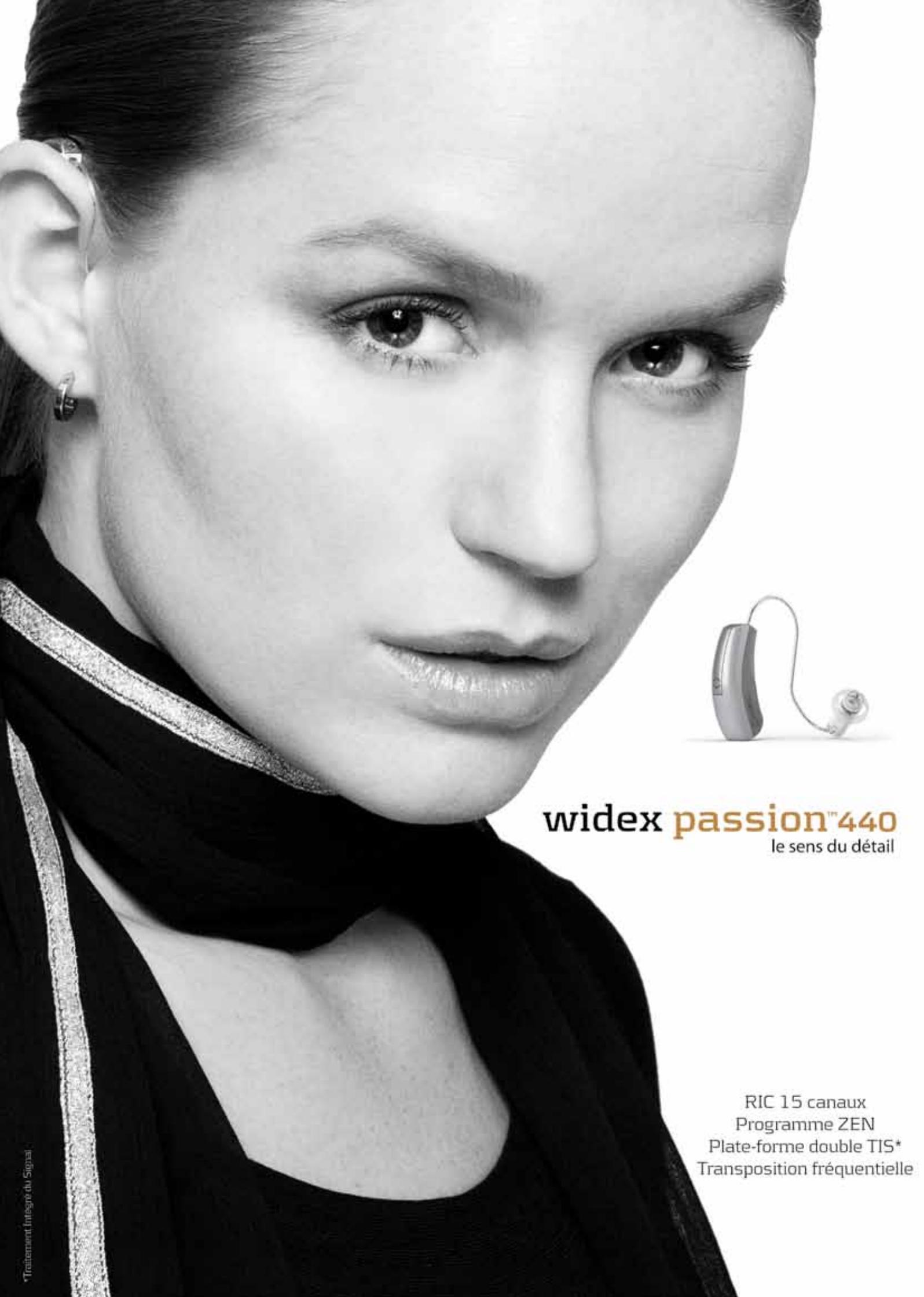


# Les Cahiers de L'AUDITION

REVUE D'INFORMATIONS TECHNIQUES ET SCIENTIFIQUES - VOL.22 - NOV./DÉC. 2009 - N°6 - ISSN 0980-3483

## DOSSIER

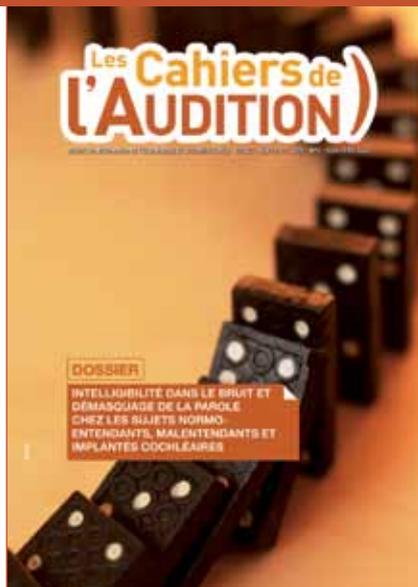
**INTELLIGIBILITÉ DANS LE BRUIT  
ET DÉMASQUAGE DE LA PAROLE  
CHEZ LES SUJETS NORMO-  
ENTENDANTS, MALENTENDANTS ET  
IMPLANTÉS COCHLÉAIRES**



**widex passion™440**  
le sens du détail

RIC 15 canaux  
Programme ZEN  
Plate-forme double TIS\*  
Transposition fréquentielle

# SOMMAIRE



## Publication de la S.A.R.L. GALATÉE

Gérant et directeur de la publication :  
Daniel CHEVILLARD - 12 ter, Rue de Bondy  
93600 Aulnay sous Bois  
Tél. : 01 48 68 19 10 - Fax : 01 48 69 77 66  
soniclaire@infonie.fr

### Rédacteur en Chef

Pr. Paul AVAN - Faculté de Médecine  
Laboratoire de Biophysique  
28, Place Henri DUNANT - BP 38  
63001 Clermont Ferrand Cedex  
Tél. : 04 73 17 81 35 - Fax : 04 73 26 88 18  
paul.avan@u-clermont1.fr

### Rédacteurs

François DEGOVE - francois.degove@wanadoo.fr  
Arnaud COEZ - acoez@noos.fr  
Assistante : C. DEGOVE  
5, avenue Maréchal JOFFRE - 92380 Garches  
Tél. 01 47 41 00 14

### Conception - Réalisation

MBQ - 32, rue du Temple - 75004 Paris  
Tél. : 01 42 78 68 21 - Fax : 01 42 71 90 83  
stephanie.bertet@mbq.fr

### Publicité

Christian RENARD - 50, rue Nationale  
BP 116 - 59027 Lille Cedex  
Tél. : 03 20 57 85 21 - Fax : 03 20 57 98 41  
contact@laborenard.fr

### Service Abonnements

Editions Elsevier Masson SAS  
62, rue Camille DESMOULINS  
92442 Issy-les-Moulineaux Cedex  
Tél. : 01 71 16 55 55 - Fax : 01 71 16 55 88  
infos@masson.fr - www.masson.fr/revues/cau

### Dépot Légal à date de parution

Nov./ Déc. 2009 - Vol. 22 - N°6  
Imprimé par Néo-typo - Besançon

N°capp 0411 T 87 860

Indexée dans : EMBASE / Excerpta Medica

## 4 Instructions aux auteurs

## 7 Éditorial

Paul AVAN

## 9 Dossier

Intelligibilité dans le bruit et démasquage de la parole  
chez les sujets normo-entendants, malentendants  
et implantés cochléaires

Dan GNANSIA

## 43 Veille technologique

Oticon  
Unitron

## 46 Informations

Compte-rendu EPU 2009  
Programme EPU 2010  
Congrès des audioprothésistes 13, 14 et 15 mars 2010  
Siemens partenaire de la Semaine du Son

**Liste des annonceurs :** Phonak • Siemens • Starkey • Widex Acourex

*Les Cahiers de l'Audition déclinent toute responsabilité sur les documents qui leur sont confiés, insérés ou non. Les articles sont publiés sous la seule responsabilité de leurs auteurs.*

**Gérant : Daniel CHEVILLARD**  
**Publicité : Christian RENARD**  
**Conception - Réalisation : MBQ**

## Rédaction

**Rédacteur en Chef :**  
 Pr. Paul AVAN

**Rédacteurs :**  
 François DEGOVE & Arnaud COEZ

## Comité de rédaction

**Audiologie Prothétique : techniques d'appareillage, d'évaluation et de contrôle de l'Adulte et de l'Enfant :**  
 Arnaud COEZ - Thierry RENGLLET

**Phonétique appliquée & audiométrie vocale :**  
 Frank LEFEVRE

**Audiologie Médicale :**  
 Docteur Jean-Louis COLLETTE  
 Professeur Paul DELTENRE

**Audiologie Expérimentale :**  
**Pr. Christian LORENZI**  
 Stéphane GALLEGO  
 Stéphane GARNIER

**Orthophonie Education et Rééducation de la Parole et du Langage :**  
 Annie DUMONT

**Veille Technologique :**  
 Robert FAGGIANO

**Veille Informatique :**

Charles ELCABACHE  
**Dossiers, Documents, Bibliographie & Communication :**  
 Bernard AZEMA  
 Arnaud COEZ  
 François DEGOVE  
 Philippe LURQUIN  
 Benoît VIROLE

## Comité associés

**Comité de Lecture : Au titre de la Société Française d'Audiologie :**

**Président :**  
 Professeur Bruno FRACHET

**Comité O.R.L. Audiophonologie :**  
**Responsable :**  
 Professeur Alain ROBIER

**Adjoints :**  
 Professeur René DAUMAN  
 Docteur Dominique DECORTE  
 Docteur Christian DEGUINE  
 Docteur Olivier DEGUINE  
 Professeur Alain DESAULTY  
 Docteur Jocelyne HELIAS  
 Docteur Jacques LEMAN  
 Docteur Lucien MOATTI  
 Docteur Jean-Claude OLIVIER  
 Docteur Françoise REUILLARD  
 Professeur François VANEECLOO  
 Professeur Christophe VINCENT

**Au titre de Présidents des Syndicats**

**Professionnels d'audioprothésistes :**  
 Patrick ARTAUD  
 Francine BERTHET  
 Frédéric BESVEL  
 Benoit ROY

**Au titre de Membres du Comité Européen des Techniques Audiologiques :**  
 Herbert BONSEL  
 Franco GANDOLFO  
 Heiner NORZ  
 Patrick VERHEYDEN

**Au titre de Directeurs de l'Enseignement de l'Audioprothèse et de D.U. :**

Professeur Lionel COLLET  
 Joël DUCOURNEAU  
 Professeur Pascale FRIANT-MICHEL  
 Professeur Alexandre GARCIA  
 Professeur Benoît GODEY  
 Professeur Jean-Luc PUEL  
 Professeur Claude SIMON  
 Professeur Patrice TRAN BA HUY

**Au titre de la Société Française d'Audiologie :**

Professeur Jean-Marie ARAN  
 Bernadette CARBONNIÈRE  
 Docteur Jean-Louis COLLETTE  
 Docteur Marie-José FRAYSSE  
 Professeur Eréa-Noël GARABEDIAN  
 Docteur Bernard MEYER  
 Docteur Martine OHRESSER  
 Docteur Sophie TRONCHE

## LES CAHIERS DE L'AUDITION SONT PUBLIÉS SOUS L'ÉGIDE SCIENTIFIQUE DU COLLÈGE NATIONAL D'AUDIOPROTHÈSE

**10, rue Molière 62220 CARVIN**  
**Tél. : 03 21 77 91 24**  
**Fax : 03 21 77 86 57**  
<http://www.college-nat-audio.fr/>

**Président :** Eric BIZAGUET

**Premier Vice-Président :**  
 Frank LEFEVRE

**Deuxième Vice-Président :**  
 Christian RENARD

**Présidents d'Honneur :**  
 Jacques DEHAUSSY  
 Xavier RENARD

**Au titre de Membres du Collège National d'Audioprothèse :**  
 Kamel ADJOUT  
 Patrick ARTHAUD  
 Jean-Claude AUDRY  
 Bernard AZEMA  
 Jean BANCONS  
 Jean-Paul BERAHA  
 Hervé BISCHOFF

Geneviève BIZAGUET  
 Jean-Jacques BLANCHET  
 Daniel CHEVILLARD  
 Arnaud COEZ  
 Christine DAGAIN  
 Ronald DE BOCK  
 Xavier DEBRUILLE  
 François DEGOVE  
 Jean-Baptiste DELANDE  
 Jean-Pierre DUPRET  
 Charles ELCABACHE  
 Robert FAGGIANO  
 Thierry GARNIER  
 Stéphane GARNIER  
 Grégory GERBAUD  
 Eric HANS  
 Bernard HUGON  
 Jérôme JILLIOT  
 Yves LASRY  
 Stéphane LAURENT  
 François LE HER  
 Maryvonne NICOT-MASSIAS

Jean OLD  
 Georges PEIX  
 Benoit ROY  
 Claude SANGUY  
 Philippe THIBAUT  
 Jean-François VESSON  
 Frédérique VIGNAULT  
 Alain VINET  
 Paul-Edouard WATERLOT

**Au titre de Membres Correspondants étrangers du Collège National d'Audioprothèse :**  
 Roberto CARLE  
 Léon DODELE  
 Philippe ESTOPPEY  
 André GRAFF  
 Bruno LUCARELLI  
 Leonardo MAGNELLI  
 Carlos MARTINEZ OSORIO  
 Thierry RENGLLET  
 Juan Martinez SAN JOSE  
 Christoph SCHWOB

## Généralités

Les travaux soumis à la rédaction des Cahiers de l'Audition sont réputés être la propriété scientifique de leurs auteurs.

Il incombe en particulier à ceux-ci de recueillir les autorisations nécessaires à la reproduction de documents protégés par un copyright.

Les textes proposés sont réputés avoir recueilli l'accord des co-auteurs éventuels et des organismes ou comités d'éthique dont ils ressortent. La rédaction n'est pas responsable des textes, dessins ou photos publiés qui engagent la seule responsabilité de leurs auteurs.

L'acceptation par la rédaction implique le transfert automatique des droits de reproduction à l'éditeur.

## Esprit de la revue

De manière générale, Les Cahiers de l'Audition sont une revue d'informations scientifiques et techniques destinée à un public diversifié : audioprothésistes, audio-logistes, orthophonistes ou logopèdes, médecins en contact avec les différents secteurs de l'audition (généralistes, neurologues, électrophysiologistes, ORL, etc...).

Ce public souhaite une information qui soit à la fois à jour sur le plan scientifique et technique, et didactique. Le but des auteurs des Cahiers de l'Audition doit être de rendre accessible cette information, même aux non-spécialistes de tel ou tel sujet.

Bien que Les Cahiers de l'Audition n'exigent pas d'un article qu'il présente des données originales, l'article lui-même doit être original, c'est-à-dire ne pas avoir déjà été publié tel quel dans une autre publication sans l'accord explicite conjoint des auteurs et de la rédaction des Cahiers de l'Audition.

## Manuscrits

Ils sont à fournir en deux exemplaires (1 original + 1 copie, complets à tous égards). La remise de manuscrits électroniques (disquettes 3 pouces 1/2, format Macintosh ou PC Word 5 ou Word 6) est vivement encouragée. Elle est destinée à l'imprimeur et ne dispense pas de l'envoi des 2 exemplaires « papier ». Ne pas faire soi-même de mise en page puisqu'elle sera faite par l'imprimeur.

Les schémas, dessins, graphiques doivent être ou des originaux ou des tirages bien contrastés, en trait noir sur papier blanc. Les tirages sur imprimante laser de qualité sont encouragés. Les diapositives de ces éléments ayant servi à une projection sont acceptées. L'encre bleue est prohibée pour des raisons techniques. Les photos doivent être de préférence des diapositives ou des tirages papier de grande qualité. Les illustrations doivent être référencées avec précision et leur emplacement souhaité dans le texte indiqué approximativement, ainsi que la taille souhaitée (noter que 1 colonne de revue = 5,5 cm de large).

**En cas de demande expresse, les documents seront retournés aux auteurs après impression.**

**Les manuscrits, rédigés en français, devront comporter en 1ère page le titre de l'article, les noms des auteurs, leurs titres, leurs adresses, une table des matières et un résumé en français et en anglais indiquant brièvement le but général de l'article, les méthodes mises en œuvre et les conclusions proposées.**

Le plan de l'article sera découpé en sections. La bibliographie ne sera pas forcément limitée à celle citée dans le texte : en effet, les auteurs peuvent rajouter quelques ouvrages de base dont ils recommandent la lecture à ceux qui souhaiteraient compléter leur information. Toutefois, l'usage extensif de références à des publications difficiles d'accès pour les lecteurs, ou trop spécialisées, n'est pas recommandé.

## Chronologie

Lorsque les auteurs ont été sollicités par un responsable de la rédaction, ils en reçoivent une confirmation écrite qui leur indique une date limite souhaitée pour la rédaction de leur article. Le respect de cette date est essentiel car il conditionne la régularité de parution de la revue. Lorsqu'un auteur soumet spontanément un article à la revue, la chronologie est indiquée ci-dessous.

Les manuscrits une fois reçus seront soumis au comité de lecture qui pourra demander des modifications ou révisions avant publication. L'avis du comité de lecture sera transmis aux auteurs dans un délai ne dépassant pas 1 mois. La publication doit donc survenir au plus tard 2 mois après réception de l'article sauf cas de force majeure (qui pourrait rajouter un délai de 3 mois). Ces indications n'ont pas valeur de contrat et le fait de soumettre un article aux Cahiers de l'Audition sous-entend l'acceptation des conditions de publication.

Une fois l'article mis en page, l'imprimeur envoie les épreuves de celui-ci à l'auteur : ces épreuves doivent être renvoyées corrigées sous 3 jours. Les seules corrections admises portent sur ce qui n'a pas été respecté par rapport au manuscrit, ou sur la mauvaise qualité de la mise en page ou de la reproduction de figures.

L'auteur ou l'équipe d'auteurs recevra 20 exemplaires gratuits du numéro de la revue où l'article est paru.

## Les manuscrits sont à adresser à

Pr. Paul AVAN

Les Cahiers de l'Audition  
Laboratoire de Biophysique  
Faculté de médecine, BP38  
63001 Clermont-Ferrand cedex, France

Il est automatique  
et si petit que vous  
l'oubliez quand  
vous le portez.



Il est si

AUDÉO **mini**™

Plus d'audition. Tout petit. Simple d'emploi.

Audéo MINI est incroyablement petit tout en étant très agréable à manier. Le tout nouvel Audéo est riche de fonctions sophistiquées, mais sa haute technologie CORE est simple d'emploi – et si bien dissimulée. Audéo MINI s'adresse à ceux qui veulent tout entendre en toute discrétion – c'est un système totalement automatique, parfaitement adapté à la vie en action.

[www.phonakpro.com](http://www.phonakpro.com)

PHONAK

life is on

## ÉDITORIAL



Le fait que le travail constituant le cœur de ce numéro des Cahiers soit issu de la thèse d'Université effectuée et soutenue par Dan Gnansia au sein d'un laboratoire de recherches prestigieux, hébergé dans un temple de la Science française, ne doit pas faire peur au lecteur. Aux antipodes d'un travail académique ésotérique et coupé de la réalité (et d'ailleurs, en existe-t-il vraiment en dehors de l'imagerie populaire ?), il est exemplaire de la littérature que les Cahiers de l'Audition cherchent à faire connaître le plus largement possible. Sa qualité pédagogique est suffisamment remarquable pour mettre à portée des débutants, et notamment des étudiants, des notions pourtant d'une redoutable complexité. Dan Gnansia avait eu l'opportunité de présenter un bref aperçu de ce travail lors de la journée scientifique du congrès de l'UNSAF 2009, et cette version est beaucoup plus complète.

La complexité du sujet traité ne peut être esquivée, elle constitue le quotidien du métier d'audioprothésiste et d'ailleurs, elle en fait tout l'intérêt. Son point de départ est une double constatation classique. Premièrement, les normoentendants arrivent à comprendre relativement facilement la parole dans un brouhaha, beaucoup mieux qu'au sein d'un bruit stationnaire : c'est le phénomène de démasquage de la parole qui repose sur l'exploitation par l'audition des 'vallées' temporelles et spectrales du bruit. Deuxièmement, les malentendants jeunes ou presbycousiques ne bénéficient pas de ce démasquage de la parole que l'appareillage est impuissant à restaurer, d'où la plainte récurrente de nos patients. Pourquoi, comment évaluer ce phénomène et y remédier, c'est l'objet de ce travail d'examiner toutes ces délicates questions. Il serait absurde d'essayer de résumer les leçons du travail de Dan Gnansia tant la lecture exhaustive de son dossier est éclairante et se suffit à elle-même. Rappelons plutôt que ce chercheur a joué une part prépondérante dans la création d'un test vocal dans le bruit à base de logatomes, rapide et simple, partie intégrante du grand programme de recherches actuel sur les bases génétiques de la presbycousie piloté par le Professeur Christine Petit. Les utilisateurs confirment tous la grande efficacité pratique de ce test qui est appelé à occuper une place de choix au sein de la panoplie actuellement disponible. Et comme pour illustrer son aisance à faire le lien entre laboratoire de recherches et terrain, Dan Gnansia travaille désormais dans l'industrie de l'appareillage auditif.

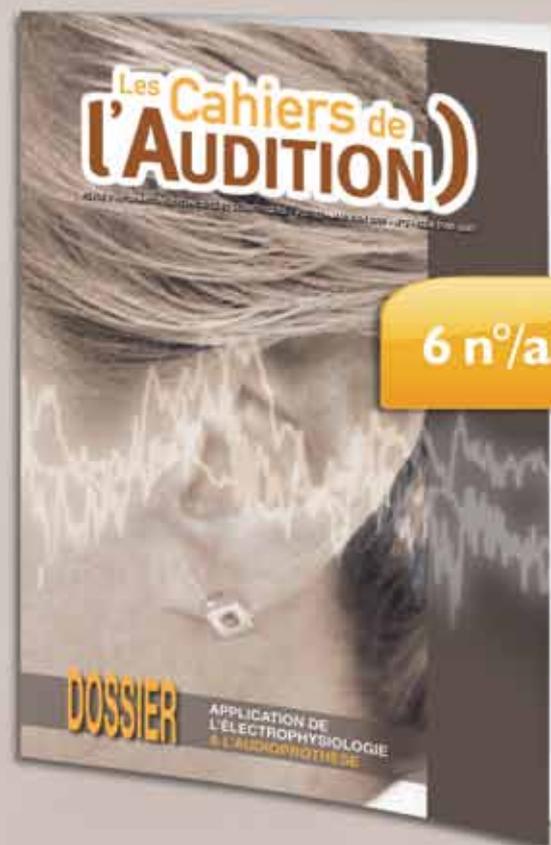
Les absents ont toujours tort, c'est ce que constateront ceux qui ne faisaient pas partie des 850 présents au dernier EPU en lisant le très synthétique compte-rendu de François Degove. Ce dernier, magnanime, leur offre la consolation d'une mise en perspective très personnelle et très pertinente, imprégnée de l'expérience qui lui permet de distinguer les directions de recherche vraiment porteuses et d'expliquer pourquoi elles le sont.

Paul Avan

# Abonnez-vous dès aujourd'hui !\*

→ LA REVUE DE TOUS LES PROFESSIONNELS DE L'AUDIOLOGIE  
SOUS L'ÉGIDE DU COLLÈGE NATIONAL D'AUDIOPROTHÈSE

- Les Cahiers de l'Audition abordent tous les sujets importants comme :
  - l'acoustique
  - la psychoacoustique
  - l'audioprothèse
  - la physiologie et la pathologie de l'oreille
  - la psychologie
  - et l'orthophonie
- Que vous soyez audioprothésiste, médecin ORL, acousticien, physiologiste, orthophoniste ou psychologue, Les Cahiers de l'Audition vous offrent un moyen exceptionnel pour être informé des évolutions de votre spécialité.



Indexation : Scopus

\*Les Cahiers de l'audition est uniquement disponible sur abonnement.

À RENVoyer À : Elsevier Masson  
Service Abonnements - 62 rue Camille-Desmoulins  
92442 Issy-les-Moulineaux cedex - www.elsevier-masson.fr

## BULLETIN D'ABONNEMENT 2010

● OUI, je souhaite m'abonner aux Cahiers de l'Audition pour un an

### COORDONNÉES

Mme  Mlle  M.

Nom \_\_\_\_\_

Prénom \_\_\_\_\_

Adresse \_\_\_\_\_

CP \_\_\_\_\_ Ville \_\_\_\_\_

e-mail \_\_\_\_\_ K10A405

J'accepte de recevoir des informations commerciales de la société Elsevier Masson par e-mail.

Conformément à la loi « Informatique et Libertés » du 6/1/1978, vous disposez d'un droit d'accès et de rectification des données personnelles vous concernant. Pour l'exercer, adressez-vous à : Elsevier Masson Service Abonnements - 62 rue Camille-Desmoulins 92442 Issy-les-Moulineaux cedex.

### AIDEZ-NOUS À MIEUX VOUS CONNAÎTRE



Votre profession : \_\_\_\_\_

Votre mode d'exercice : \_\_\_\_\_

### TARIFS TTC FRANCE 2010

<input type="checkbox"/> Etudiant*	<input type="checkbox"/> Particulier	<input type="checkbox"/> Institution
46 €	91 €	139 €

\* Sur justificatif

Pour les tarifs à l'étranger, veuillez contacter notre Service Abonnements : 01 71 16 55 99 - info@elsevier-masson.fr  
Votre abonnement commence avec le n°1 - 2010. Pour tout bulletin d'abonnement reçu après le 30 octobre 2010, nous mettrons en place un abonnement 2011 ; si vous souhaitez maintenir un abonnement 2010, cochez cette case :

### RÈGLEMENT

Ci-joint mon règlement d'un montant de : \_\_\_\_\_ € TTC

Par chèque bancaire ou postal à l'ordre d'Elsevier Masson

Par carte bancaire :  Visa  Eurocard/MasterCard

N° \_\_\_\_\_

Cryptogramme visuel

(3 derniers n° au dos de votre CB) :

N° \_\_\_\_\_

Date d'expiration : \_\_\_\_\_

Date et signature obligatoires :

# INTELLIGIBILITÉ DANS LE BRUIT ET DÉMASQUAGE DE LA PAROLE CHEZ LES SUJETS NORMO-ENTENDANTS, MALENTENDANTS ET IMPLANTÉS COCHLÉAIRES

## Introduction

Une musique de fond sur un dialogue au cinéma, le brouhaha des clients du bar pendant une conversation au comptoir, la rumeur de la rue alors que l'on discute sur un banc public, ou tout simplement les instruments accompagnant les paroles d'une chanson...

Autant de bruits de fond, de sons concurrents et de situations quotidiennes d'écoute de la parole. Pourtant, force est de constater que pour la plupart des personnes jeunes ne souffrant pas de troubles auditifs ou langagiers, les capacités de compréhension orale n'en sont pas moins conservées dans l'ensemble de ces situations. Un brouhaha – en raison de ses fluctuations spectrales et temporelles – peut même s'avérer moins gênant (moins masquant) qu'un bruit de fond stationnaire tel que le bruit produit, par exemple, par un ventilateur ou un climatiseur. Ce phénomène remarquable et peut être surprenant en première approche semble toutefois d'une grande complexité (par le nombre de mécanismes auditifs impliqués) et d'une grande fragilité. En effet chez les personnes âgées ou souffrant de pertes auditives même légères, ce phénomène qualifié du nom de « *démasquage de la parole* » diminue voire disparaît, conduisant à l'isolement complet de ces personnes dans la plupart des situations quotidiennes (*e.g.*, repas familiaux, télévision, restaurant, etc). Très clairement, dans tous ces moments présentant du bruit de fond, et où la grande majorité des personnes jeunes et entendant comprennent chaque mot prononcé, la plupart des personnes malentendantes, jeunes ou âgées, ne comprennent strictement rien. Par ailleurs, bien que les dispositifs de réhabilitation auditive actuels (prothèses auditives et implants cochléaires) semblent maintenant restituer une bonne intelligibilité dans le silence, les personnes malentendantes

appareillées ou implantées sont d'une façon générale en très grande difficulté dans le bruit, et ne présentent pas plus de capacité de démasquage qu'avant appareillage.

Comment mesurer avec précision les capacités de démasquage de la parole ? Quels facteurs influencent ce phénomène ? Quels mécanismes sont impliqués dans cette capacité ? Quelles sont les causes de la réduction du démasquage chez les personnes malentendantes ou implantées ? Que proposer pour les nouvelles prothèses et implants cochléaires afin de rétablir le démasquage de la parole chez les personnes malentendantes ? Nous allons tenter d'apporter quelques réponses à ces questions énoncées pour certaines il y a près de soixante ans, mais toujours d'une grande actualité en sciences de l'audition de la parole et en audiologie.

## CHAPITRE 1 INTELLIGIBILITÉ EN PRÉSENCE DE MASQUES NON STATIONNAIRES ET DÉMASQUAGE DE LA PAROLE

1

Définition et caractérisation empirique

Étude princeps : Miller et Licklider (1950)

Dans une étude considérée aujourd'hui comme princeps, Miller et Licklider (1950) rapportent des mesures systématiques de nos capacités à identifier des signaux de parole – l'intelligibilité de la parole – dans plusieurs conditions de distorsion sévère. Différentes conditions expérimentales sont testées dans cette étude : signal de parole seul et

Adapté de la thèse de doctorat de l'Université Pierre et Marie Curie

Présentée par Dan Gnansia<sup>1,2,3,4</sup>

Sous la direction de :  
Alain de Cheveigné<sup>2,3</sup>  
et Christian Lorenzi<sup>2,3,4</sup>

- 1 : Société Neurelec, France  
2720 Chemin St Bernard  
06224 Vallauris Cedex, France
- 2 : Laboratoire de Psychologie de la Perception - UMR CNRS 8158, Université Paris Descartes and École Normale Supérieure, 45 rue des Saint Pères, 75006 Paris, France
- 3 : Département d'Études Cognitives, École Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France
- 4 : GDR CNRS 2967 GRAEC École Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

interrompu périodiquement, signal de parole continu en présence d'un bruit masquant stationnaire ou d'un bruit interrompu périodiquement, et parole et bruit interrompus périodiquement et alternativement. Les stimuli sont de simples mots monosyllabiques.

L'un des résultats majeurs de cette étude est que l'intelligibilité (les performances d'identification correcte des signaux de parole) s'avère *bien meilleure* en présence de bruit interrompu que de bruit stationnaire. Cette amélioration varie certes en fonction du niveau de présentation du masque, de la fréquence et du rapport cyclique des interruptions, mais illustre avant tout la très grande **robustesse** ou **résistance** de notre système de traitement de la parole face à la corruption de la parole produite par les signaux concurrents généralement **non stationnaires** de notre environnement (le brouhaha produit par un ou plusieurs convives pendant un repas ou un cocktail, par exemple).

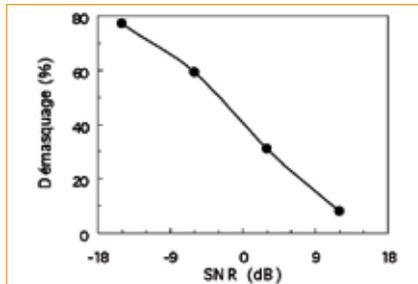


FIGURE 1.1 : NIVEAU DE DÉMASQUAGE DE LA PAROLE EXPRIMÉ EN TERMES DE DIFFÉRENCE DE POURCENTAGE D'INTELLIGIBILITÉ (% CORRECT EN BRUIT FLUCTUANT - % CORRECT EN BRUIT STATIONNAIRE) EN FONCTION DU RAPPORT SIGNAL-SUR-BRUIT À LONG TERME (SNR EN DB). D'APRÈS MILLER ET LICKLIDER (1950).

La figure 1.1 résume les résultats de cette étude, en présentant la quantité de « démasquage » (masking release), autrement dit la différence entre performance mesurée en bruit fluctuant et performance mesurée en bruit stationnaire, en fonction du rapport signal-sur-bruit à long terme, pour un bruit interrompu à 10 Hz avec un rapport cyclique de 50 %. Point important, le niveau de référence en condition de bruit stationnaire correspond à l'énergie à long terme du bruit interrompu pour permettre le calcul du démasquage.

Ces résultats fondateurs sur le phénomène de démasquage de la parole mettent en lumière un mécanisme clé (ou plusieurs mécanismes clés, comme nous le verrons plus loin) de l'intelligibilité de la parole en situation écologique. En effet, un auditeur normo-entendant semble pouvoir *tirer profit des fluctuations temporelles ou spectrales d'un signal masquant*, notamment des minima d'énergie par bande de fréquence de ce masque, pour récupérer des fragments d'information phonétique (c'est-à-dire des informations partielles sur le signal de parole) et ainsi améliorer l'intelligibilité de la parole, par rapport au cas où le masque est stationnaire et de même énergie à long terme.

Nous verrons par la suite 1) quels sont les paramètres physiques du signal de parole contribuant à ce/ces mécanisme(s) remarquables, 2) quels types de mécanismes sensoriels ou cognitifs sont supposés être impliqués dans ce phénomène, mais aussi 3) que ce dernier est relativement fragile car fortement affecté par des pertes auditives neurosensorielles parfois légères, et 4) difficilement compensable par prothèse conventionnelle ou implant cochléaire.

L'utilisation de bruit stationnaire que l'on module pour y introduire des fluctuations temporelles ou spectrales correspond à un besoin de contrôle expérimental pour l'étude de la perception de la parole dans un bruit fluctuant. En effet, ce type de masque permet d'une part une étude quantitative et précise du démasquage dans certaines conditions choisies car les caractéristiques acoustiques des vallées (les minima) spectro-temporelles du bruit (régularité d'apparition, durée, forme, profondeur, etc.) sont parfaitement maîtrisées, et d'autre part de déterminer les principales caractéristiques acoustiques qui rendent possible l'extraction d'un son cible dans le mélange parole/bruit. Toutefois, nous allons voir par la suite que d'autres masques peuvent être utilisés, et notamment d'autres signaux de parole concurrents.

## 2

### Effets paramétriques

#### a. Effet du nombre de locuteurs interférents

Afin de se rapprocher d'un contexte de masquage écologique de type « Cocktail Party » (Cherry, 1953), de nombreuses études sur l'intelligibilité de la parole ont été menées en utilisant des signaux de parole concurrents (natural babble) pour masquer le signal de parole cible.

Les études de Miller (1947) présentent les premières mesures d'intelligibilité de la parole en fonction du nombre de locuteurs interférents. La variable dépendante utilisée dans cette étude est le SRT (Speech Reception Threshold) qui correspond à la valeur en décibels du rapport signal-sur-bruit nécessaire à l'obtention d'une performance d'identification correcte de 50 % (ou 40 %, variable selon les études). Miller rapporte une augmentation de 8 dB SRT entre 1 et 2 locuteurs interférents, et une autre augmentation de 3 à 4 dB SRT entre 2 et 4, 6 et 8 locuteurs interférents. En conclusion de cette étude, plus le nombre de signaux de parole concurrents est grand, plus le masquage de la parole (la dégradation de l'intelligibilité) est important.

Pourtant, l'intelligibilité de la parole est meilleure lorsque la cible est masquée par 4 à 8 locuteurs interférents qu'avec un masque en comportant une infinité (Duquesnoy, 1983 ; Festen and Plomp, 1990). Ce bruit stationnaire ayant le même spectre de puissance à long terme que la parole, il est par convention dénommé bruit « Speech Shaped Noise » (SSN). Ces résultats révèlent un aspect important du ou des mécanismes de démasquage de la parole, car un nombre limité de locuteurs interférents produit un masque dont les fluctuations (temporelles et spectrales) sont auditivement très saillantes, c'est-à-dire détectables et discriminables par les sujets normo-entendants. De la même façon, Danhauer et Lepper (1979) ont relevé de meilleurs scores d'intelligibilité lorsque la parole est masquée par 4 à 9 locuteurs que lorsqu'elle est masquée par un bruit stationnaire de même énergie à long terme (pour une revue de l'ensemble de ces résultats, voir Bronkhorst, 2000).

L'utilisation de locuteurs interférents comme masque est certes écologique, mais met inéluctablement en branle des processus de traitement de nature spécifiquement linguistique, et de facto certains mécanismes impliqués dans le masquage dit informationnel (lié ici au contenu sémantique du masque) que nous détaillerons par la suite. C'est pour limiter au maximum le masquage à ses composantes les plus sensorielles que plusieurs études sur l'intelligibilité de la parole dans le bruit utilisent un masque fluctuant construit à partir d'un bruit SSN, modulé par l'enveloppe des signaux de parole interférents. Un tel bruit est souvent appelé « *babble-modulated noise* » (e.g., Festen and Plomp, 1990 ; Bronkhorst and Plomp, 1992 ; Peters *et al.*, 1998 ; Versfeld and Dreschler, 2002 ; Cooke, 2006). Ainsi, le contenu sémantique du masque est sévèrement limité car ce dernier n'est plus ou peu intelligible, mais les fluctuations temporelles restent similaires à celles présentes au sein du signal de parole interférent d'origine.

L'étude de Cooke (2006) utilise ce type de masque afin de valider un modèle de démasquage de la parole que nous détaillerons dans une section ultérieure. Cette étude tente en effet de prédire l'intelligibilité de la parole dans un masque de type *babble-modulated noise*, et ce pour plusieurs valeurs du nombre de locuteurs interférents. Les résultats expérimentaux initiaux sont rapportés dans une étude réalisée par Simpson et Cooke (2005).

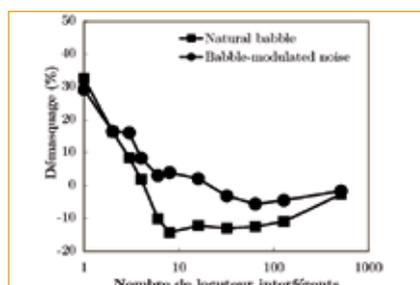


FIGURE 1.2 : POURCENTAGE D'IDENTIFICATION DE CONSONNES (VCV, VOYELLE/CONSONNE/VOYELLE) EN FONCTION DU NOMBRE DE LOCUTEURS INTERFÉRENTS (N) POUR UN « NATURAL BABBLE » (TRIANGLES, TRAIT PLEIN), ET POUR UN « BABBLE-MODULATED NOISE » (RONDS, TRAIT POINTILLÉ). D'APRÈS SIMPSON ET COOKE (2005).

La figure 1.2 reprend les résultats de Simpson et Cooke (2005), à savoir l'évolution du pourcentage d'identification de

non-mots bisyllabiques (VCV, Voyelle/Consonne/Voyelle) en fonction du nombre de locuteurs interférents pour un natural babble et pour un babble-modulated noise. Des différences d'intelligibilité notables apparaissent entre les deux types de bruit lorsque le nombre de locuteurs interférents est compris entre 4 et 128, suggérant fortement que le natural babble véhicule certaines informations acoustiques et/ou linguistiques – absentes du babble-modulated noise – accentuant le masquage de la parole.

## b. Effets du bruit modulé en amplitude

Dans plusieurs études plus contrôlées que les précédentes (mais aux dépense de l'« écologie » de la situation), le masque utilisé est un simple bruit large bande modulé en amplitude. Le babble-modulated noise, que nous venons de voir, est un masque de ce type, mais il présente des minima temporels peu réguliers et un spectre de modulation complexe (riche en composantes de modulation spectre d'amplitude variable). Ainsi, afin de déterminer – entre autres – l'effet de la durée et de la profondeur de ces instants pendant lesquels le bruit est moins énergétique et le démasquage de la parole rendu possible, des modulations plus simples et régulières telles que des modulations sinusoïdales ou carrées ont été appliquées (des signaux dont le spectre de modulation se résume à une simple raie ou à une série de composantes connues dont les caractéristiques sont aisément manipulables). Le choix du bruit porteur est le plus souvent un bruit SSN (plutôt qu'un bruit blanc) car un tel bruit produit un effet de masquage audio comparable sur l'ensemble des bandes de fréquence typiques de la parole (Horii *et al.*, 1970; Berry and Nerbonne, 1972).

Voyons maintenant les effets des paramètres physiques de la modulation du masque sur le démasquage de la parole.

### Effet de la fréquence de modulation :

L'étude de Miller et Licklider (1950) discutée en introduction de ce chapitre nous donne un premier aperçu de l'effet

de la fréquence de modulation (et donc de la durée des instants de silence au sein du bruit) sur le démasquage de la parole. En effet, Miller et Licklider rapportent que l'intelligibilité de phrases présentées dans un masque à modulations carrées est maximale pour des fréquences comprises entre environ 2 et 10 Hz, cette valeur pic variant selon le rapport signal-sur-bruit de présentation (pour -18 dB, intelligibilité maximum à 8 Hz ; pour 9 dB, intelligibilité maximum à 15 Hz). Pour les valeurs supérieures, le démasquage diminue jusqu'à être quasiment aboli pour des fréquences de modulation supérieures à 100 Hz, très probablement en raison des limitations introduites par l'acuité temporelle du système auditif humain. Fait étonnant, les auteurs observent une diminution du démasquage pour les faibles valeurs de fréquence de modulation. La fonction caractérisant le démasquage dans le domaine des modulations d'amplitude aurait donc une allure sélective (passe bande).

Des ordres de grandeur similaires ont été constatés par Gustafsson et Arlinger (1994). Sur des stimuli de type phrase, ces auteurs relèvent un pic de démasquage de 45 points de pourcentage quand le masque est modulé à des fréquences comprises entre 10 et 20 Hz, puis une diminution jusqu'à abolition du démasquage à partir de 100 Hz de fréquence de modulation. Par contre, contrairement à Miller et Licklider, Gustafsson et Arlinger n'observent pas de diminution nette du démasquage pour les faibles valeurs de fréquence de modulation.

Avec des stimuli composés de non-mots bisyllabiques (VCV, Voyelle/Consonne/Voyelle), plusieurs auteurs remarquent un maximum de démasquage de 35 points de pourcentage pour un masque modulé sinusoïdalement à 8 Hz (Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006b), avec une diminution marquée pour les plus faibles et plus hautes valeurs de fréquence de modulation, accompagnée d'une quasi abolition du démasquage à partir de 100 Hz. L'utilisation de stimuli VCV offre certains avantages comme celui de permettre de calculer la quantité d'information reçue pour certains traits phonétiques (Miller and Nicely, 1955) tels le voisement, le

lieu d'articulation et le mode d'articulation perçu. Les maxima de démasquage observés pour ces trois traits phonétiques sont différents : le démasquage exprimé en terme de voisement reçu est maximal pour des fréquences de modulation allant de 2 à 64 Hz, le lieu d'articulation présente un pic à 32 Hz chez les sujets normo-entendants jeunes (Füllgrabe *et al.*, 2006) et à 8 Hz chez les sujets normo-entendants âgés (Lorenzi *et al.*, 2006b), et le mode d'articulation montre un pic de démasquage pour une fréquence de modulation de 8 Hz.

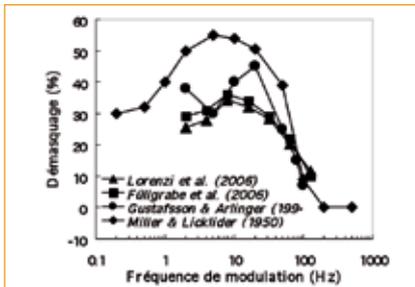


FIGURE 1.3 : DÉMASQUAGE DE LA PAROLE EXPRIMÉ EN TERMES DE DIFFÉRENCE DE POURCENTAGE, EN FONCTION DE LA FRÉQUENCE DE MODULATION DU MASQUE. D'APRÈS LORENZI ET AL. (2006B), RÉALISÉ AVEC STIMULI VCV (TRIANGLES) ; FÜLLGRABE ET AL. (2006) RÉALISÉ AVEC STIMULI VCV (CARRÉS) ; GUSTAFSSON ET ARLINGER (1994) RÉALISÉ AVEC STIMULI PHRASES (CERCLES) ; ET MILLER ET LICKLIDER (1950) RÉALISÉ AVEC STIMULI MOTS (LOSANGES).

La figure 1.3 résume les résultats de ces quatre études en présentant le démasquage observé en fonction de la fréquence de modulation du masque. Comme indiqué plus haut, un maximum de démasquage est généralement observé entre 5 et 32 Hz environ. Ce trait semble commun à la plupart des études. Les variations – parfois importantes – observées néanmoins entre études semblent être issues des différences de protocoles et de stimuli (et de formes de modulation notamment). En effet, Miller et Licklider (1950) utilisent des rapports signaux sur bruit constants à travers les sujets (pour la figure 1.3, le rapport signal-sur-bruit à long terme est de -6 dB), alors que les autres études utilisent un rapport signal-sur-bruit réglé individuellement pour lequel le sujet atteint 30 % (Gustafsson and Arlinger, 1994) ou 50 % (Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006b) de réponses correctes dans la condition de bruit stationnaire. De plus, Miller et Licklider modulent le bruit à l'aide d'un signal carré, alors que Füllgrabe *et al.* et

Lorenzi *et al.* modulent ce bruit à l'aide d'une sinusoïde, tandis que Gustafsson et Arlinger avec un signal périodique complexe. Toutefois, la comparaison de ces différents résultats permet d'affirmer que le démasquage de la parole semble optimal pour des fréquences de modulation comprises entre 5 et 32 Hz, ce qui correspond à une fenêtre temporelle optimale de 16 à 100 ms, variable selon le type d'indice acoustique ou phonétique extrait ou traité par le sujet. De nombreuses autres études confirment ces tendances (*e.g.*, Licklider and Guttman, 1957 ; Festen, 1987 ; Kwon and Turner, 2001 ; Dubno *et al.*, 2002 ; Buss *et al.*, 2003 ; Dubno *et al.*, 2003 ; Nelson *et al.*, 2003 ; Rhebergen *et al.*, 2006). Le démasquage de la parole impliquerait donc une analyse auditive des mélanges parole/bruit fluctuant à différentes échelles de temps, analyse probablement contrainte tant par la structure temporelle des événements articulatoire-acoustiques du signal de parole (structure syllabique, explosions d'amplitude, temps d'attaque vocale, transitions formantiques, etc.) que par l'acuité temporelle du système auditif.

#### Effet de la profondeur de modulation :

Plusieurs études évoquent le rôle de la profondeur de modulation du masque (et du rapport signal-sur-bruit « local » dans les vallées du bruit) dans le démasquage de la parole, mais seuls Howard-Jones et Rosen (1993a) ont tenté d'en mesurer les effets de façon systématique. En effet, Howard-Jones et Rosen utilisent un masque fluctuant produit par une modulation carrée de fréquence 10 Hz, dont ils font varier la profondeur de modulation. À l'aide de stimuli VCV, ces auteurs mesurent l'intelligibilité en fonction du niveau du bruit pendant les minima, induisant ainsi un changement du rapport signal-sur-bruit à long terme (les maxima d'amplitude sont maintenus constants).

Les auteurs remarquent que *plus le niveau de bruit dans les minima est faible (plus la vallée temporelle du bruit est profonde), meilleure est l'intelligibilité de la parole*. Cette amélioration semble d'ailleurs évoluer de manière linéaire pour le protocole utilisé et la variable dépendante choisie. Conformé-

ment à cette étude, d'autres recherches rapportent une meilleure intelligibilité pour des profondeurs de modulation du bruit élevées (*e.g.*, Gustafsson and Arlinger, 1994 ; George *et al.*, 2006).

Récemment, Gnansia *et al.* (2008) ont étudié les effets de la profondeur de modulation du bruit masquant sur le démasquage de la parole. Dans cette étude, les performances d'identification de consonnes au sein de syllabes sans signification sont mesurées chez des auditeurs normo-entendants, en présence d'un bruit SSN stationnaire et d'un bruit SSN modulé sinusoïdalement en amplitude à 8 et 32 Hz. Ici, les performances d'identification sont mesurées à différentes profondeurs de modulation (de 0 à 100 %) du bruit SSN, avec un rapport signal-sur-bruit fixé individuellement afin de produire environ 50 % d'intelligibilité en condition de bruit stationnaire. Conformément aux résultats discutés plus haut, les mesures indiquent que le démasquage augmente en fonction de la profondeur de modulation du bruit (voir figure 1.4). Par ailleurs, un démasquage significatif est observé dès les faibles valeurs de profondeur de modulation du bruit, correspondant à un rapport signal-sur-bruit local dans la vallée du bruit de -5 dB.

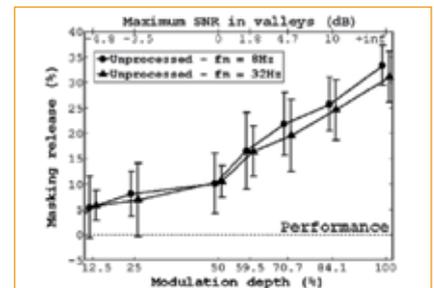


FIGURE 1.4 : DÉMASQUAGE DE LA PAROLE EN FONCTION DE LA PROFONDEUR DE MODULATION DU BRUIT. LE BRUIT MODULÉ PAR UN SINUS DE FRÉQUENCE 8 (RONDS) OU 32 Hz (TRIANGLES). LA PROFONDEUR DE MODULATION EST EXPRIMÉE EN POURCENTAGE (ABSCISSE DU BAS) ET EN RAPPORT SIGNAL-SUR-BRUIT LOCAL DANS LES VALLÉES DU BRUIT (ABSCISSE DU HAUT). D'APRÈS GNANSIA ET AL. (2008).

Ces études convergent donc vers l'idée que les sujets « entr'aperçoivent » (glimpse ; voir Cooke and Ellis, 2001 ; Cooke, 2003) le signal de parole (c'est-à-dire extraient des informations partielles de parole) dans les minima du bruit de fond, lorsque la durée et la profondeur de ces minima le permettent

(lorsqu'elles sont supérieures à une valeur seuil donnée, contrainte perceptivement par l'acuité temporelle et fréquentielle du système auditif). Dans deux études basées sur un modèle informatique « d'écoute dans les vallées des bruit » qui sera détaillé plus loin, Cooke (2006) tente de déterminer ces valeurs seuils. Ainsi, la profondeur minimale de la vallée des bruits conduisant par simulation du modèle à des performances de démasquage similaires à celles des sujets normo-entendants correspond à un rapport signal-sur-bruit local de -5 ou -2 dB (en fonction de paramètres du modèle de reconnaissance de parole). Ainsi, conformément aux résultats de Gnansia *et al.* (2008), la profondeur minimale d'une vallée de bruit engendrant du démasquage (exprimée en rapport signal-sur-bruit local dans la vallée du bruit) est de l'ordre de -5 dB.

Ici, il est important de noter que ces valeurs sont obtenues pour des stimuli VCV et un babble-modulated noise. Cette étude aurait-elle conduit aux mêmes résultats si elle avait été menée à l'aide d'un bruit modulé par un signal périodique comme c'est le cas dans plusieurs études antérieures ? D'autres travaux semblent donc nécessaires afin de s'en assurer et mieux caractériser la capacité « d'écoute dans les vallées ».

### Effet de la forme de la modulation :

Les travaux portant sur l'effet de la fréquence et de la profondeur de modulation du masque utilisent des formes de modulation différentes. En effet, certains utilisent des modulations carrées (*e.g.*, Miller and Licklider, 1950 ; Howard-Jones and Rosen, 1993a ; Dubno *et al.*, 2002 ; 2003 ; Nelson *et al.*, 2003 ; George *et al.*, 2006 ; Rhebergen *et al.*, 2006), sinusoïdales (*e.g.*, Takahashi and Bacon, 1992 ; Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006b ; Rhebergen *et al.*, 2006), ou périodiques et complexes (*e.g.*, Gustafsson and Arlinger, 1994 ; Füllgrabe *et al.*, 2006 ; Rhebergen *et al.*, 2006). D'après ce que nous venons de conclure sur la profondeur de modulation, plus la vallée de bruit est profonde, plus le phénomène de « glimpsing » est important, et ainsi, meilleure est l'intelligibilité (et donc le démasquage). On peut vérifier en **figure 1.3** que le plus

fort effet de démasquage est observé dans l'étude de Miller et Licklider (1950), cette dernière utilisant un masque à modulations carrées, c'est-à-dire un masque dont les vallées sont bien mieux définies (les pentes d'enveloppe étant plus abruptes) que pour une modulation sinusoïdale par exemple.

Ceci se vérifie également en **figure 1.5** : Rhebergen *et al.* (2006) ont ici comparé les SRT pour l'identification de phrases dans un bruit fluctuant présentant des modulations carrées ou sinusoïdales, ce à plusieurs fréquences de modulation. Pour les fréquences de modulation les plus basses, les interruptions carrées conduisent à une meilleure intelligibilité que les modulations sinusoïdales. Ces résultats suggèrent une nouvelle fois que l'intelligibilité est augmentée lorsque les vallées du bruit sont mieux définies.

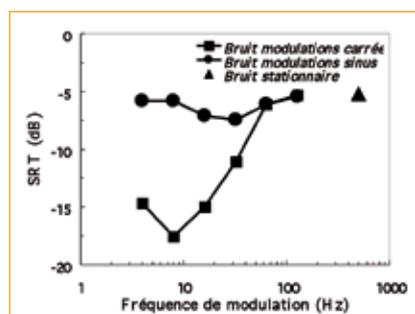


FIGURE 1.5 : SRT EN FONCTION DE LA FRÉQUENCE DE MODULATION DU BRUIT MASQUANT, POUR UN BRUIT MODULÉ PAR UN SIGNAL CARRÉ (CARRÉS), UNE SINUSOÏDE (CERCLES), OU SANS MODULATION (TRIANGLE), DANS UNE TÂCHE D'IDENTIFICATION DE PHRASES. D'APRÈS RHEBERGEN ET AL. (2006).

Les travaux de Füllgrabe *et al.* (2006) fournissent également des informations quant aux effets de la forme des modulations du bruit masquant. Dans cette étude, Füllgrabe *et al.* comparent les valeurs de démasquage de la parole obtenues sur des sujets normo-entendants jeunes pour deux types de modulation du bruit masquant. Les fluctuations du masque sont obtenues en appliquant une modulation sinusoïdale à l'amplitude du bruit dans un premier temps (modulation sinusoïdale dite du premier ordre), puis à la profondeur de modulation d'un bruit déjà modulé sinusoïdalement en amplitude (modulation sinusoïdale dite du second ordre). Les résultats indiquent que le démasquage

est indépendant de la fréquence de la variation de la profondeur, suggérant d'une part que les caractéristiques de la distribution temporelle des vallées du masque n'influent pas notablement sur les résultats, et d'autre part que c'est bien la profondeur et la durée des vallées du masque qui sont les facteurs cruciaux du phénomène de démasquage.

### Autres effets :

D'autres paramètres, comme le rapport cyclique, peuvent influencer le démasquage de la parole. Plusieurs travaux rapportent les effets de ce paramètre sur le démasquage de la parole (*e.g.*, Nelson *et al.*, 2003 ; Rhebergen *et al.*, 2006) : le démasquage diminue à mesure que le rapport cyclique augmente, à rapport signal-sur-bruit moyen constant. Une fois de plus, plus la vallée de bruit est conséquente, meilleure est l'intelligibilité dans le bruit fluctuant, et donc le démasquage.

Le choix des stimuli et la méthode d'évaluation sont bien entendu des facteurs déterminants de la mesure d'intelligibilité et de démasquage. Ainsi, les études diffèrent selon qu'il s'agit de l'identification de syllabes sans signification (*e.g.*, Howard-Jones and Rosen, 1993a ; b ; Assmann and Summerfield, 1994 ; Dubno *et al.*, 2002 ; 2003 ; Buss *et al.*, 2004 ; Cutler *et al.*, 2004 ; Cooke, 2006 ; Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006b), de mots (*e.g.*, Miller and Licklider, 1950 ; Carhart *et al.*, 1975 ; Snell *et al.*, 2002 ; Buss *et al.*, 2003 ; Turner *et al.*, 2004), ou de phrases (*e.g.*, Plomp and Mimpen, 1979 ; Duquesnoy, 1983 ; Festen and Plomp, 1990 ; Gustafsson and Arlinger, 1994 ; Dorman *et al.*, 1998 ; Nelson and Jin, 2004 ; George *et al.*, 2006). Ces stimuli diffèrent considérablement par leur complexité à plusieurs niveaux linguistiques (acoustique, phonétique, phonologique, morphologique, syntaxique et sémantique). Segmentation et/ou accès au lexique sont requis dans certains cas (*e.g.*, reconnaissance de mots ou phrases) alors que seule l'identification de segments phonétiques isolés est requise dans d'autres (*e.g.*, reconnaissance de logatomes sans signification). Par ailleurs, la mesure d'intelligibilité

peut être réalisée à partir d'une tâche en choix forcé parmi un nombre limité d'alternatives (le sujet choisit un non-mot/ un mot/une phrase au sein d'une liste fermée) ou non. Enfin, la métrique choisie pour la mesure de l'intelligibilité peut être différente selon l'étude : il peut s'agir du nombre de réponses correctes, mais aussi du nombre de mots-clé correctement identifiés dans une phrase (e.g., Nelson and Jin, 2004), ou encore du nombre de mots correctement identifiés dans une phrase (e.g., Dorman *et al.*, 1998 ; George *et al.*, 2006). Enfin, pour la mesure de l'intelligibilité dans le bruit et du démasquage de la parole, la grandeur choisie peut être soit un pourcentage de réponse correcte, soit un niveau de bruit pour lequel le sujet obtient 40 ou 50 % de réponses correctes (le SRT), soit une différence de pourcentage entre la performance obtenue avec un bruit fluctuant et celle obtenue avec un bruit stationnaire.

Ces variations importantes entre paradigmes expérimentaux limitent parfois fortement la possibilité de comparer les résultats des nombreuses études portant sur le démasquage, et expliquent pour partie certaines divergences notées quant aux effets des paramètres étudiés ci-dessus.

### c. Effets du bruit modulé spectralement

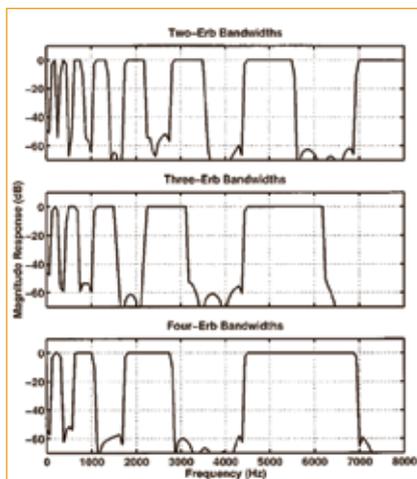


FIGURE 1.6 : ENVELOPPE SPECTRALE DE LA MODULATION APPLIQUÉE AU BRUIT MASQUANT, RÉALISANT DES VALLÉES SPECTRALES D'UNE PÉRIODICITÉ DE 2 ERB<sup>N</sup> (HAUT), 3 ERB<sup>N</sup> (MILIEU), OU 4 ERB<sup>N</sup> (BAS). D'APRÈS PETERS ET AL. (1998).

Comme suggéré par Cooke (2003 ;

2006), les sujets normo-entendants tirent profit de minima spectraux et temporels locaux du bruit masquant pour améliorer l'intelligibilité de la parole. Nous avons considéré les effets des modulations temporelles du masque, et allons maintenant envisager le cas des modulations spectrales.

Une étude de Peters *et al.* (1998) examine les effets d'une modulation temporelle et/ou spectrale sur le démasquage de la parole, chez des sujets normo-entendants et malentendants. Pour réaliser les modulations spectrales, les auteurs filtrent le bruit masquant stationnaire (de type SSN) dans plusieurs régions fréquentielles. Le principe consiste à utiliser un banc de filtres passe-bande de bande passante égale à 2, 3 ou 4 ERB<sup>N</sup>, et dont les fréquences centrales sont uniformément réparties sur une échelle ERB<sup>N</sup> (1/2, 1/3 ou 1/4 de filtre par ERB<sup>N</sup>), puis de mettre à zéro la sortie d'un filtre sur deux (périodiquement dans le domaine fréquentiel). La **figure 1.6** illustre cette périodicité de 2, 3 ou 4 ERB<sup>N</sup> introduite dans les bruits masquants. La modulation spectro-temporelle – si désirée – est obtenue en appliquant le filtrage décrit ci-dessus à un bruit préalablement modulé temporellement par l'enveloppe du signal de parole d'un locuteur interférent.

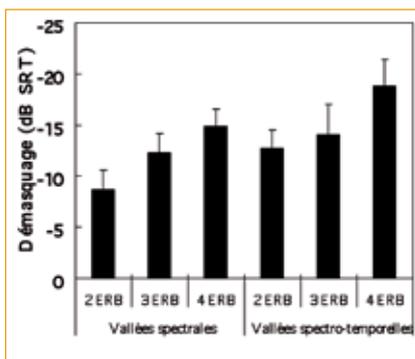


FIGURE 1.7 : DÉMASQUAGE DE LA PAROLE (EN DB SRT) POUR UN GROUPE DE 10 SUJETS NORMO-ENTENDANTS, POUR PLUSIEURS CONDITIONS DE BRUIT (MODULATIONS SPECTRALES ET SPECTRO-TEMPORELLES, PLUSIEURS VALEURS EN ERB<sup>N</sup> DES VALLÉES SPECTRALES) DANS UNE TÂCHE D'IDENTIFICATION DE PHRASES. D'APRÈS PETERS ET AL. (1998).

Les mesures d'intelligibilité obtenues sur un groupe de 10 sujets normo-entendants sont présentées en **figure 1.7** en termes de démasquage de la parole en dB SRT (différence entre score obtenu en bruit fluctuant et score obtenu en

bruit stationnaire), pour chaque condition expérimentale. Tout d'abord, un démasquage est observé dès la plus faible largeur de vallée spectrale (-9 dB SRT pour 2 ERB<sup>N</sup>), révélant ainsi l'efficacité de ce mécanisme d'écoute dans les vallées spectrales. Par ailleurs, plus la largeur des vallées de bruit augmente, meilleur est le démasquage. Enfin, les meilleurs scores de démasquage sont observés dans les conditions de modulation spectro-temporelles du bruit, suggérant que les sujets normo-entendants combinent les vallées spectrales et temporelles pour mieux démasquer la parole.

D'autres études confirment cette capacité de combinaison des vallées spectrales et temporelles. Par exemple, Howard-Jones et Rosen (1993b) démontrent à l'aide de masques présentant des vallées spectrales et temporelles déphasées (masque dit « en damier ») que, dans une certaine mesure, les sujets normo-entendants peuvent regrouper des informations dans différentes régions fréquentielles et à différents instants pour démasquer au mieux la parole (capacité appelée « *uncomodulated glimpsing* » par les auteurs). De la même façon, en utilisant des masques modulés temporellement et spectralement, Buss *et al.* (2003 ; 2004) constatent qu'un démasquage est observable pour des vallées spectrales ou spectro-temporelles asynchrones au sein d'un bruit.

Pris ensemble, ces résultats suggèrent donc que les sujets entendants présentent une capacité 1) à *extraire des portions d'informations phonétiques* dans des vallées du bruit limitées temporellement et spectralement et ne présentant pas de synchronie ni d'organisation spectrale ou temporelle particulière, et 2) à *combinaison ces segments de parole* pour améliorer l'intelligibilité.

## 3

### Mécanismes du démasquage

#### a. Nature des indices auditifs

D'un point de vue acoustique, la parole est un signal présentant une structure spectrale et temporelle complexe.

L'identification de ces formes complexes peut être conçue comme reposant d'une part sur la perception d'**indices auditifs spectraux** fournis par la décomposition fréquentielle réalisée par la membrane basilaire (aboutissant à une représentation spatiale ou tonotopique de ces indices), et d'autre part sur la perception d'**indices auditifs temporels** fournis par les distributions de décharges neurales des fibres afférentes du nerf auditif. La distinction entre indices spectraux et temporels n'est toutefois pas aisée car les indices fréquentiels dits de « place d'excitation » (*i.e.*, tonotopiques) peuvent être à la fois statiques (formants des voyelles) et dynamiques (transitions formantiques caractérisant l'articulation entre deux phonèmes). Les indices temporels peuvent être conçus de manière plus restrictive comme correspondant aux fluctuations de l'amplitude ou de la fréquence instantanée du signal de parole au sein de chaque bande de fréquence cochléaire (*i.e.*, pour chaque place d'excitation dans la cochlée). Ces fluctuations temporelles peuvent alors être divisées sommairement – peut être avec un certain degré d'arbitraire – en trois groupes en fonction de leurs corrélats acoustiques, perceptifs et linguistiques (voir **figure 1.8**), comme cela a été proposé dans une étude descriptive extrêmement influente réalisée par Rosen (1992) : les fluctuations les plus lentes (entre 2 et 50 Hz environ) qui correspondent aux informations transmises par *l'enveloppe d'amplitude* du signal au sens strict (véhiculant notamment les informations sur l'intensité, la durée, le temps d'attaque et de chute du signal de parole, le tempo), et les fluctuations plus rapides (supérieures à environ 50 Hz) qui correspondent à la *périodicité* et à la *structure temporelle* fine du signal. Les fluctuations de périodicité comprises entre 50 et 800 Hz environ véhiculent en particulier l'information sur la fréquence fondamentale du signal (F0), ou fondamentale laryngée (Terhardt, 1972a ; b), mais aussi le voisement et la qualité de la voix (Rosen, 1992), alors que les fluctuations de structure temporelle fine semblent véhiculer essentiellement des informations de timbre, de lieu d'articulation, de voisement et certaines informations sur le mode d'articulation (contraste nasales/non nasales).

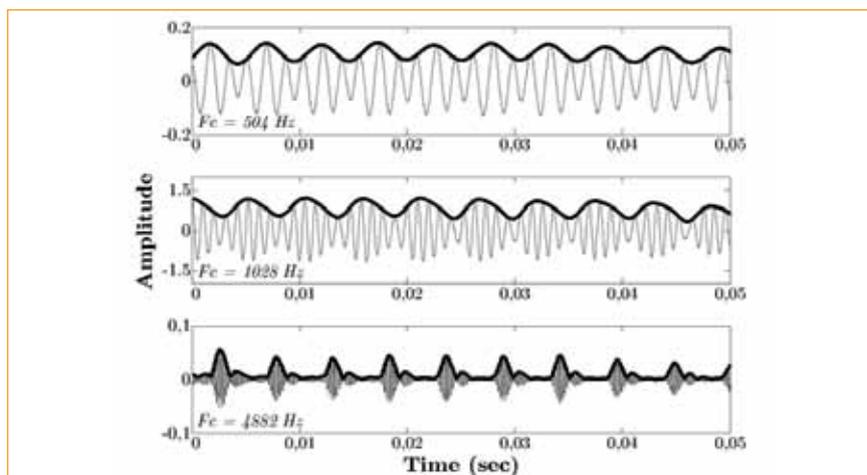


FIGURE 1.8 : SIGNAL D'ONDE EN SORTIE DE FILTRES COCHLÉAIRES SIMULÉS CENTRÉS SUR 504, 1028 ET 4882 HZ, RÉPONDANT AU SON « A » DE « ABA » EN FRANÇAIS. LA LIGNE ÉPAISSE ET LA LIGNE FINE MONTRENT RESPECTIVEMENT LE RÉSULTAT D'UNE DÉCOMPOSITION DES SIGNAUX EN BANDE ÉTROITE EN UNE ENVELOPPE D'AMPLITUDE (MODULATEUR DE LA PORTEUSE) ET UNE STRUCTURE TEMPORELLE FINE (PORTEUSE).

Ces indices temporels peuvent également, et toujours probablement avec un certain degré d'arbitraire, être regroupés en deux catégories, résultant d'une décomposition du signal à bande étroite en chaque point de la membrane basilaire en une enveloppe temporelle et une structure temporelle fine. En première approche et du point de vue du signal, l'enveloppe temporelle et la structure temporelle fine peuvent être conçues respectivement comme un modulateur et un signal porteur (le module et l'argument de la transformée de Hilbert du signal sonore en bande étroite), mais du point de vue de la cochlée, comme les éléments du codage spatial et temporel. En effet, l'enveloppe temporelle correspondant aux fluctuations lentes et énergétiques du signal peut être considérée comme le principal élément du codage tonotopique le long de la membrane basilaire. La structure temporelle fine, suivant quant à elle les fluctuations rapides de la fréquence instantanée du signal, correspond aux informations temporelles proprement dites dans le nerf auditif, notamment grâce aux phénomènes de verrouillage de phase des fibres du nerf auditif (décharges dans le nerf auditif synchronisées avec la structure temporelle fine jusqu'à environ 5 kHz ; Johnson, 1980). Cette **dichotomie enveloppe/structure temporelle fine** fournira le cadre théorique dans lequel cette recherche doctorale a été conduite. Il est toutefois important de noter que cette dichotomie, bien que partiellement fondée

sur un mode de décomposition possible des signaux de parole et sur certaines données psychoacoustiques, demande à être plus amplement démontrée sur le plan neurophysiologique.

Ces fluctuations lentes et rapides, autrement dites d'enveloppe temporelle et de structure temporelle fine, semblent jouer un rôle capital dans l'intelligibilité de la parole. En effet, l'enveloppe temporelle même dégradée semble suffisante pour une excellente intelligibilité (*e.g.*, Shannon *et al.*, 1995 ; Smith *et al.*, 2002). Ce constat est conforme aux résultats de plusieurs études pionnières portant sur la perception d'enveloppe et l'intelligibilité dans le silence sous implant cochléaire, suggérant que la réception d'information phonétique dépend de manière critique de la qualité de la réception des informations d'enveloppe temporelle lente et rapide (*e.g.*, Cazals *et al.*, 1994). Concernant l'intelligibilité dans le bruit et le démasquage de la parole, l'importance qualitative et quantitative de l'enveloppe temporelle (liée également à la résolution spectrale) et de la structure temporelle fine sera précisée par la suite. On peut cependant indiquer en première approche que, si l'enveloppe temporelle au sein d'un nombre limité de bandes de fréquence (4-8 bandes) suffit pour identifier correctement un signal de parole dans le silence, la structure temporelle fine semble jouer un rôle majeur dans le bruit (*e.g.*, Nelson *et al.*, 2003 ; Qin and Oxenham, 2003 ; Hopkins *et al.*, 2008).

**Widex Equipement,**  
l'expertise technique pour bien choisir.

**Présence – Ecoute – Conseil**

**Multi-marques**

**Installation**

**Expertise informatique**

**Chaînes de mesure**

**Audiomètres**

**Impédancemètres**

**Oto-émissions acoustiques**

**PEA et PEA automatiques**

**Réparation**

**Etalonnage**

**Stock**

**Matériel de prêt**



Depuis 1980, Widex fournit, installe et répare une gamme complète d'équipements audiolgiques, à la pointe de la technologie pour les professionnels de la santé.

## b. Ecoute dans les vallées du bruit

Le démasquage est plus important pour un bruit masquant présentant des modulations sinusoïdales que pour des modulations reprenant l'enveloppe temporelle de la parole (Bacon *et al.*, 1998). Nous avons vu qu'il augmente avec la profondeur de modulation du bruit modulé en amplitude (Howard-Jones and Rosen, 1993a ; Gustafsson and Arlinger, 1994) et les fréquences de modulation optimales pour observer ce phénomène se situent entre 8 et 25 Hz (Miller and Licklider, 1950 ; Gustafsson and Arlinger, 1994 ; Kwon and Turner, 2001 ; Nelson *et al.*, 2003 ; Füllgrabe *et al.*, 2006) ceci dépendant du type de signal de parole utilisé. De plus, le démasquage augmente lorsque l'on introduit des vallées spectrales dans un bruit stationnaire ou modulé temporellement en amplitude, particulièrement lorsque la largeur de ces vallées spectrales devient plus importante (Peters *et al.*, 1998). Par ailleurs, le démasquage est faible pour un bruit masquant de type « locuteurs multiples » présentant une enveloppe spectrale et temporelle relativement plate, et se réduit encore plus si les portions les plus faibles de la parole présentée sont masquées par un bruit stationnaire (Eisenberg *et al.*, 1995 ; Bacon *et al.*, 1998). Enfin, comme nous l'avons indiqué dans une section précédente (l.2.c.), un démasquage est observé en présence de bruit présentant des vallées spectrales ou spectro-temporelles synchrones ou asynchrones (Howard-Jones and Rosen, 1993b ; Buss *et al.*, 2003 ; 2004).

Ces résultats démontrent que les sujets sont capables de tirer profit des minima – parfois extrêmement circonscrits sur le plan spectro-temporel – au sein du bruit afin de détecter des informations partielles de parole. Cette capacité est souvent appelée « **écoute dans les vallées** » (GLIMPSING) : *les sujets normo-entendants peuvent regrouper des informations dans différentes régions fréquentielles et à différents instants* pour démasquer au mieux la parole.

De fait, tous ces mécanismes nécessitent un certain niveau de résolution temporelle (c'est-à-dire une capacité à suivre les fluctuations du masque afin

d'extraire de l'information pendant les vallées) et spectrale (avoir accès à des portions du signal de parole qui ne sont pas ou peu masquées dans le domaine spectral). En rapport avec cette notion, une baisse du démasquage est notée pour des fréquences de modulation du bruit masquant supérieures à 30 Hz (des vallées de bruit de moins de 17 ms), cette baisse pouvant être mise sur le compte d'effets de masquage proactif, un facteur limitant la résolution temporelle auditive (Festen, 1993 ; Dubno *et al.*, 2002). Toutefois, il est important de pointer que la démonstration d'une stratégie d'écoute dans les vallées du bruit n'explique pas entièrement nos capacités de démasquage. En effet, une autre question reste à élucider, à savoir comment le système auditif s'y prend-il (quels indices auditifs sont utilisés par l'auditeur) pour savoir que des informations de parole se trouvent dans les vallées du bruit ?

## c. Démasquage de co-modulation

La présence de modulation au sein du masque est susceptible d'améliorer la perception de la parole masquée à plus d'un titre. En effet, plusieurs études psychoacoustiques indiquent l'existence d'un phénomène de « démasquage de co-modulation » (CMR ou comodulation masking release) reposant sur un traitement auditif des fluctuations du masque à travers les bandes de fréquence audio (Hall *et al.*, 1984 ; Hall and Grose, 1988 ; Hall *et al.*, 1988). L'effet CMR correspond au fait que la détection d'un son pur dans un bruit fluctuant (à bande étroite autour du son pur) est facilitée par l'adjonction d'un autre bruit co-modulé (présentant des modulations cohérentes avec celles du premier bruit) dans une plage de fréquence éloignée.

Le rôle du CMR dans la perception de la parole dans le bruit a été initialement étudié par Grose et Hall (1992). Afin de se replacer dans les conditions des études psychophysiques précédentes, Grose et Hall utilisent des signaux de parole (voyelles) passés par trois filtres passe bande étroits (de bande passante 40 Hz) à trois fréquences centrales (445, 890 et 1780 Hz), masqués par un bruit

filtré soit de façon similaire, soit passé par douze filtres de fréquence centrales différentes (incluant les trois bandes de filtrage du signal de parole). Les résultats montrent que la co-modulation du bruit masquant n'affecte pas l'intelligibilité des voyelles. Ainsi, si le CMR joue un rôle dans les tâches de détection, celui-ci ne semble pas avoir d'effet dans les tâches d'identification des signaux vocaliques.

Une étude basée sur une méthode comparable (stimuli de type consonne, filtrés passe bande) montre que l'adjonction d'une bande de bruit co-modulée dans une plage de fréquence éloignée du signal cible n'améliore pas le démasquage de la parole. Toutefois, cette adjonction ne le dégrade pas, contrairement à ce qui est observé lorsque cette bande de bruit supplémentaire n'est pas co-modulée (Kwon, 2002).

Dans une autre étude, Festen (1993) utilise un paradigme plus sophistiqué où des phrases sont masquées par une voix concurrente dont les modulations d'amplitude par bande de fréquence sont co-modulées ou non (par simple décalage temporel entre bandes de fréquence). Conformément aux autres études, les résultats de l'étude de Festen indiquent que la contribution du CMR à l'intelligibilité dans le bruit et au démasquage de la parole est relativement faible, ne produisant une amélioration que de 1dB SRT. Cet effet modeste et parfois difficile à démontrer est néanmoins à prendre en compte pour une pleine compréhension du démasquage de la parole.

## d. Ségrégation parole/masque et groupement auditif

Certaines études suggèrent que les mécanismes de « groupement auditif » impliqués dans *l'analyse des scènes auditives* jouent un rôle dans le démasquage de la parole (*e.g.*, Bregman, 1990). En effet, la ségrégation auditive du signal de parole cible et du masque en *flux auditifs* (auditory streams ; représentations auditives associées à des sources acoustiques distinctes) devrait être facilitée par l'introduction de disparités d'enveloppe temporelle entre signal et bruit et par la co-modulation du bruit, et conduire ainsi à une augmentation de l'intelligibilité dans le bruit.

D'après Bregman (1990), *l'analyse de la scène auditive* fait appel à deux grandes classes de mécanismes de ségrégation, à savoir la ségrégation simultanée et séquentielle, selon que les événements sonores se recouvrent dans le temps ou non. Chacun de ces mécanismes est, comme nous allons le voir, susceptible de participer au démasquage de la parole.

### Ségrégation simultanée :

Ces mécanismes de ségrégation ont été largement étudiés grâce au paradigme dit des doubles voyelles (voir de Cheveigné, 1999, pour une revue). Outre le rapport des intensités, les deux paramètres principaux contrôlant cette forme de ségrégation auditive sont 1) la différence de fréquence fondamentale  $f_0$  (e.g., Scheffers, 1983 ; Zwicker, 1984 ; Assmann and Summerfield, 1990 ; Stickney *et al.*, 2004) et son évolution dans le temps (Binns and Culling, 2007) entre les voyelles, et 2) la synchronisation des attaques (onsets) et des chutes d'amplitude (offsets, voir par exemple Darwin, 1984). Ce deuxième type d'indice est un indice d'enveloppe temporelle. Il est d'évidence introduit par les fluctuations les plus lentes de l'amplitude du masque dans les situations de démasquage de la parole décrites plus haut (e.g., lorsque le bruit est modulé par un signal carré), et pourrait donc contribuer à l'amélioration de l'intelligibilité. Néanmoins, force est de constater que les rares études portant sur ce point ne rapportent qu'une faible corrélation entre SRT mesuré avec de la parole concurrente et ségrégation simultanée (Summers and Leek, 1998).

### Ségrégation séquentielle :

Le rôle de la ségrégation séquentielle dans le démasquage semble plus clair au vu de la littérature, et mérite de fait d'être plus amplement détaillé. Les premiers travaux sur la ségrégation séquentielle sont probablement ceux réalisés par Miller et Heise (1950). Ces auteurs observent que deux sons répétés alternativement (séquence ABABA...) produisent des percepts différents : les sons semblent venir d'une seule source ou de deux sources incohérentes selon

l'écart de fréquence entre les deux sons. Ce phénomène d'organisation (de ségrégation) séquentielle est plus communément appelé « *streaming* » et signifie que des événements sonores temporellement disjoints peuvent être regroupés dans un même flux auditif.

Ce phénomène d'organisation auditive a par la suite été précisé par VAN NOORDEN (1975) en utilisant une séquence de son légèrement différente (séquence ABA – ABA – ..., un rythme de galop est perçu quand la séquence fusionne auditivement en un seul flux). En demandant au sujet d'essayer d'entendre le rythme de galop (1 seul flux) aussi longtemps qu'il le peut alors que la différence de fréquence entre les sons A et B augmente, VAN NOORDEN observe l'existence d'un seuil au-delà duquel la perception d'un seul flux devient impossible. Ce seuil est appelé seuil de cohérence temporelle et reflète une situation où la ségrégation est irrépessible. Si, au contraire, VAN NOORDEN demande au sujet d'essayer d'entendre le rythme régulier formé par la suite de sons A, tout en réduisant la différence de fréquence qui sépare les deux types de sons, il obtient un autre seuil au delà duquel il est impossible d'entendre deux flux. Ce seuil est appelé le seuil de fission et reflète une situation où la ségrégation est volontaire. VAN NOORDEN (1975) a exploré ces deux seuils pour différents tempos. Ces résultats sont illustrés en **figure 1.9**. La dépendance ou non d'un phénomène de streaming vis-à-vis du tempo permet d'évaluer si la mesure concerne le seuil de cohérence temporelle ou le seuil de fission.

À l'aide d'une méthode psychophysique similaire, Bregman (1978) met en évidence une phase de « construction » ou d'« accumulation » (build-up) dans l'effet de streaming : pour que les deux flux se forment, le sujet doit être exposé à la séquence de sons pendant environ 4 s. Ce temps, relativement long comparé aux durées caractéristiques des processus cognitifs, serait nécessaire à l'accumulation d'indices auditifs permettant la construction des flux. De la même façon, un certain temps semble nécessaire à la déconstruction des flux : le fait d'interrompre le signal pendant quelques secondes ne supprime pas entièrement l'effet de streaming.

Ainsi, quand le signal reprend, il est possible d'observer un biais sur l'effet de ségrégation dépendant de la durée du silence.

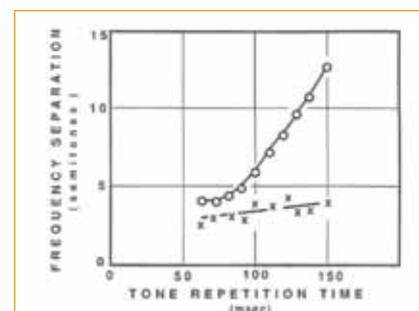


FIGURE 1.9 : SEUIL DE COHÉRENCE TEMPORRELLE (CERCLES) ET SEUIL DE FISSON (CROIX) EN FONCTION DU TEMPO (TONE REPETITION TIME). EN DESSOUS DU SEUIL DE FISSON, UN SEUL FLUX EST TOUJOURS PERÇU. AU-DESSUS DU SEUIL DE COHÉRENCE TEMPORRELLE, DEUX FLUX SONT TOUJOURS PERÇUS. ENTRE CES DEUX SEUILS, IL Y A INCERTITUDE, ET DONC BISTABILITÉ. D'APRÈS VAN NOORDEN (1975) REPRODUIT PAR BREGMAN (1990).

Dans le cas des signaux de parole, on peut noter que plusieurs indices influencent l'organisation séquentielle. Ainsi, des différences de structure formantique (Thomas *et al.*, 1970 ; Lackner and Goldstein, 1974 ; Dorman *et al.*, 1975), de timbre (Singh, 1987 ; Bregman *et al.*, 1990), ou de fréquence fondamentale (Broadbent and Ladefoged, 1957 ; Darwin and Bethell-Fox, 1977 ; Nootboom *et al.*, 1978 ; Iyer *et al.*, 2007) sont autant de facteurs pouvant affecter les seuils de cohérence et de fission (pour une évolution récente de ce champ d'étude, voir aussi Pressnitzer and Hupé, 2006).

Dans le cas des situations de démasquage décrites plus haut, (par exemple lorsque deux phrases sont prononcées simultanément) on se trouve dans une situation intermédiaire où la plupart des événements sonores ne sont ni réellement simultanés ni réellement séquentiels. Les deux mécanismes de ségrégation, simultanée et séquentielle, sont donc vraisemblablement impliqués dans la ségrégation de voix concurrentes et dans le démasquage, et interagissent certainement.

Cependant, les études menées jusqu'à maintenant sur ces mécanismes de ségrégation sont encore loin de faire le lien entre l'analyse de la scène auditive et le démasquage de la parole (pour une

revue des liens entre analyse de scène auditive et intelligibilité de la parole masquée, voir Cooke and Ellis, 2001 ; Gaudrain, 2008 ; ainsi que Gaudrain *et al.*, 2008). On notera néanmoins une étude réalisée par Mackersie *et al.* (2001) sur des sujets normo-entendants et malentendants rapportant une corrélation significative et importante entre mesures de ségrégation séquentielle sur sons purs (mesure de seuil de fission) et intelligibilité de la parole en présence de signaux de parole concurrents. Plusieurs études portant sur le démasquage de la parole (*e.g.*, Nelson *et al.*, 2003 ; Qin and Oxenham, 2003 ; Zeng *et al.*, 2005) suggèrent également que la disparition du phénomène de démasquage observée lorsque la résolution fréquentielle ou les informations de structure temporelle fine sont dégradées artificiellement par des systèmes de synthèse de parole serait due à une altération de l'organisation séquentielle. Les indices spectraux et de structure temporelle fine influenceraient donc l'écoute dans les vallées du bruit en permettant de ségréger la parole du bruit.

Ces éléments suggèrent donc que le ou les mécanismes d'écoute dans les vallées du bruit impliqués dans le démasquage ne sont pas indépendants des mécanismes d'analyse de scène auditive et de ségrégation tels que décrits par Bregman et van Noorden.

### e. Restauration phonémique

Des phénomènes de plus haut niveau tels que celui de restauration phonémique sont susceptibles d'influencer le démasquage de la parole. En effet, l'écoute dans les vallées des bruits ne fournit à l'auditeur que des informations de parole partielles et temporellement non adjacentes. Plusieurs études (Warren, 1970 ; Iyer *et al.*, 2007) révèlent toutefois que le système auditif peut percevoir les phonèmes manquants en utilisant la redondance des stimuli vocaux aux niveaux acoustique, phonétique, phonologique et surtout lexical. Ainsi, un mot ou une portion de mot dont la présence est hautement probable dans le contexte d'une phrase peut être « entendu » même s'il est acoustiquement dégradé ou absent. Dans une

étude considérée comme princeps, Warren (1970) démontre que si l'on remplace complètement un fragment de parole par un son étranger (un bruit de toux par exemple) au sein d'une phrase, le segment de parole manquant est malgré tout perçu. Cet effet n'est cependant observé que dans certaines conditions : le fragment de parole manquant est entendu à condition que le son qui le remplace soit relativement intense, et que son contenu spectral soit proche de celui du segment de parole manquant.

En lien avec ce phénomène de restauration phonémique, d'autres études indiquent que l'identification de signaux de parole par des sujets normo-entendants est très robuste vis-à-vis d'interruptions périodiques par des silences (*e.g.*, Miller and Licklider, 1950 ; Powers and Speaks, 1973 ; Powers and Wilcox, 1977 ; Nelson and Jin, 2004 ; Iyer *et al.*, 2007). La **figure 1.10** reprend les résultats de trois de ces études : l'identification de signaux de parole interrompus est maximale et quasi parfaite pour des interruptions périodiques de cadences comprises entre 8 et 50 Hz, et minimale pour les fréquences d'interruption inférieures à 2 Hz et supérieures à 100 Hz.

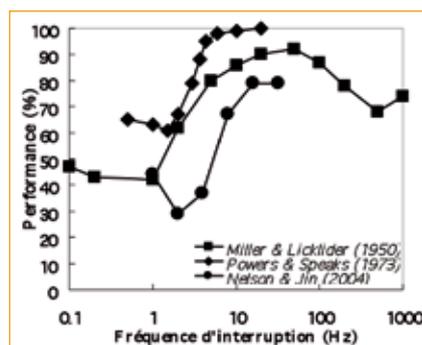


FIGURE 1.10 : POURCENTAGE D'IDENTIFICATION EN FONCTION DE LA FRÉQUENCE D'INTERRUPTION, AVEC DES STIMULI DE TYPE PHRASE. D'APRÈS MILLER ET LICKLIDER (1950), POWERS ET SPEAKS (1973), ET NELSON ET JIN (2004).

Pris ensemble, ces résultats suggèrent donc que la capacité de reconstruire la parole, ou, autrement dit, à déduire la structure spectro-temporelle complète de la parole à partir d'informations incomplètes, joue un rôle important dans le démasquage de la parole.

### f. Masquage de modulation

Le démasquage de la parole correspond à une amélioration de l'intelligibilité de la parole en présence d'un bruit masquant modulé temporellement en amplitude par rapport au cas où le bruit est stationnaire. Cependant, dans certaines conditions, il est possible d'observer le scénario inverse : l'intelligibilité de la parole peut également être réduite par les modulations du bruit, en raison d'interférences dans le traitement auditif entre les modulations temporelles du masque et celles de la parole (Kwon and Turner, 2001).

Ces effets d'interférence dans le domaine de l'intelligibilité de la parole sont à rapprocher des résultats d'expériences psychoacoustiques (menées sur signaux non-linguistiques) de mesure de la sensibilité à la modulation d'amplitude d'un signal en présence d'un masque modulé. La détection d'une modulation d'amplitude appliquée à un signal est réduite (jusqu'à 10 dB) lorsque le masque est également modulé. Cet effet est observé pour des signaux porteurs sinusoïdaux (*e.g.*, Yost and Sheft, 1989 ; Yost *et al.*, 1989 ; Bacon and Konrad, 1993 ; Bacon and Moore, 1993 ; Yost and Sheft, 1994 ; Bacon *et al.*, 1995 ; Strickland and Viemeister, 1996), et des bruits large bande (*e.g.*, Bacon and Grantham, 1989 ; Houtgast, 1989 ; Ewert and Dau, 2000 ; Millman *et al.*, 2002). Le premier phénomène est usuellement appelé *interférence de détection de modulation (modulation detection interference ou MDI)* et s'observe lorsque les signaux porteurs des modulations cible et masque sont très éloignés sur l'axe des fréquences. L'effet MDI révèle donc l'existence de processus de traitement inter-canaux (à l'instar de l'effet CMR décrit plus haut). Le second phénomène est usuellement appelé *masquage de modulation*, et s'observe lorsque les modulations cible et masque sont appliquées à la même porteuse. Le masquage de modulation révèle donc l'existence de processus de traitement spécifiques des modulations au sein de chaque canal fréquentiel audio. MDI et masquage de modulation présentent toutefois des caractéristiques similaires, ils sont en effet dépendants de la fréquence de la modulation du signal et du masque :

plus ces fréquences de modulation sont proches, plus les effets d'interférence et de masquage sont importants. Ces effets de masquage suggèrent donc l'existence d'une sélectivité fréquentielle dans le domaine des modulations analogue à la sélectivité fréquentielle audio (Dau *et al.*, 1997).

Nous avons vu plus haut que les modulations d'amplitude (ou enveloppe temporelle) jouent un rôle important dans l'intelligibilité de la parole. L'enveloppe de la parole véhicule en effet des informations phonétiques critiques, notamment lorsque la sélectivité fréquentielle audio est limitée (Van Tasell *et al.*, 1987 ; Shannon *et al.*, 1995). De plus, les modulations d'amplitude comprises entre 4 et 16 Hz semblent jouer un rôle capital pour l'identification de la parole. Ainsi, une dégradation des fluctuations d'enveloppe temporelle (par masquage ou filtrage, par exemple) n'affecte pas l'intelligibilité tant que les modulations comprises entre 4 et 16 Hz ne sont pas affectées (Drullman *et al.*, 1994a ; b).

Par conséquent, les modulations temporelles du masque présentent potentiellement deux effets antagonistes : elles permettent une ségrégation auditive parole/bruit et une écoute de la parole dans les vallées du bruit, mais elles peuvent parallèlement engendrer des effets d'interférence et de masquage de modulation susceptibles de gêner la perception des modulations de la parole cible et de réduire l'amplitude du phénomène de démasquage de la parole. Cependant, les caractéristiques spécifiques de l'effet MDI ne permettent pas de prédire ces interférences perceptives dans le domaine de la parole. En effet, les effets délétères des modulations masquées diminuent à mesure que la parole et le masque fluctuant se rapprochent dans le domaine fréquentiel (Kwon and Turner, 2001). Ceci suggère que les mécanismes supposés être impliqués dans le masquage de modulation et le MDI sont différents de ceux impliqués dans le démasquage de la parole. Certains auteurs supposent que les fluctuations du masque causent en fait une sorte d'illusion perceptive affectant l'attention du sujet : ce dernier traite alors le masque comme une partie du signal de parole (Kwon and Turner, 2001 ; Nelson *et al.*, 2003).

## g. Masquage informationnel

Un dernier phénomène connu sous le nom de *masquage informationnel* est susceptible d'influencer le démasquage de la parole (pour une revue sur le masquage informationnel, voir Durlach *et al.*, 2003). En effet, de nombreuses études suggèrent que la similarité entre les locuteurs cibles et interférents peut dégrader l'intelligibilité de la parole dans des situations de parole concurrente : plus grande est cette similarité, plus faible est l'intelligibilité du signal cible. Conformément à ceci, le choix d'un locuteur interférent différent de la cible du point de vue du genre (voix d'homme versus voix de femme), ou de l'intelligibilité (langue maternelle versus étrangère) a un effet important sur l'intelligibilité, bien que ce choix n'ait pas d'impact sur le masquage énergétique à proprement parler. Cet excès de masquage est souvent appelé masquage informationnel (Bronkhorst, 2000 ; Drullman and Bronkhorst, 2000 ; Brungart, 2001 ; Brungart *et al.*, 2001).

Une façon de réduire ce masquage informationnel (ou tout au moins, sa composante linguistique) consiste à rendre le masque inintelligible. Nous avons vu précédemment que dans de nombreuses études, le masque choisi est un bruit inintelligible de type « Babble modulated noise » (cf. 1.2.a.). Cependant, même si de tels masques présentent les mêmes fluctuations d'enveloppe que la parole, la structure spectrale et les fluctuations temporelles rapides sont fortement dégradées.

Une autre solution permettant de limiter l'intelligibilité du masque consiste à utiliser un ou plusieurs locuteurs interférents et à présenter le signal inversé temporellement (Festen and Plomp, 1990 ; Summers and Molis, 2004 ; Rhebergen *et al.*, 2005). Ainsi, le spectre de puissance du signal de parole interférent est conservé. Les résultats montrent par ailleurs que le démasquage est bien meilleur quand le masque est inversé temporellement.

Cependant, cette amélioration du démasquage par inversion temporelle du masque ne semble pas être causée uniquement par une réduction du masquage informationnel. En effet, inverser temporellement la parole

altère les fluctuations d'enveloppe (leur phase), et donc la façon dont le signal cible est masqué. Pour chaque segment, l'attaque devient la chute, ce qui inverse les effets de masquage proactif et rétroactif. Pour tester cette hypothèse, Rhebergen *et al.* (2006) utilise un locuteur interférent parlant une langue étrangère, ce afin de limiter l'intelligibilité du masque sans modifier les caractéristiques temporelles et fréquentielles « naturelles » de la parole. En comparant l'effet de masque produit par la même langue et une langue étrangère, chaque masque étant inversé temporellement ou non, Rhebergen *et al.* estiment que la contribution du masquage informationnel est substantielle, et de l'ordre de 6 à 7 dB SRT.

## 4

### Démasquage et atteintes neurosensorielles

#### a. Surdités cochléaires légères, modérées-sévères

##### Déficits associés à une perte auditive d'origine cochléaire

Les dommages subis par la cochlée sont à l'origine de nombreux déficits perceptifs. Ainsi, la détérioration des cellules ciliées externes entraîne la perte des processus actifs d'amplification non-linéaire, et engendre à la fois élévation des seuils absolus, recrutement de sonie et perte de sélectivité fréquentielle. Les dommages causés aux cellules ciliées internes engendrent une élévation supplémentaire des seuils auditifs absolus et certains troubles de la transduction, particulièrement dans le domaine temporel (pour une revue, voir Moore, 2007). Chacun de ces déficits est susceptible d'affecter le démasquage de la parole qui, comme nous le verrons plus loin, est fortement réduit – voire aboli – chez les sujets atteints de pertes auditives neurosensorielles.

##### Audibilité : Élévation des seuils absolus d'audition

Les pertes auditives neurosensorielles sont généralement associées à une élévation des seuils absolus d'audition, mesurée en dB HL (pour « Hearing



# PRÉCIS D'AUDIOPROTHÈSE

Production, phonétique acoustique et perception de la parole  
Édité par ELSEVIER MASSON SAS - ISBN N° 978-2-294-06342-8

**NOUVEAUTÉ**

## LOGICIEL LA CIBLE MÉTHODES DE CHOIX PROTHÉTIQUE

**Pré-réglage** : Xavier Renard - Membre du CNA

**CTM** : François Le Her - Membre du CNA

Production : Collège National d'Audioprothèse

Réalisation : Audition France Innovation



### Précis d'audioprothèse - Production, phonétique acoustique et perception de la parole

99,00 € x .....	exemplaire(s)	= .....	€
+ frais de port France :	8,50 € x .....	= .....	€
+ frais de port Etranger :	10,00 € x .....	= .....	€

### La Cible - Méthodes de Choix Prothétique - Pré-Réglage - CTM

150,00 € x .....	exemplaire(s)	= .....	€
+ frais de port France :	3,50 € x .....	= .....	€
+ frais de port Etranger :	4,50 € x .....	= .....	€

**SOIT UN RÈGLEMENT TOTAL** .....€

**(exonéré de TVA)**

NOM..... PRÉNOM.....

SOCIÉTÉ .....

ADRESSE .....

.....

CODE POSTAL..... VILLE .....

TÉL..... FAX .....

E-MAIL .....

### BON DE COMMANDE À ENVOYER AVEC VOTRE CHÈQUE À :

Collège National d'Audioprothèse - 10 rue Molière - 62220 CARVIN

Tél 03 21 77 91 24 - [College.Nat.Audio@orange.fr](mailto:College.Nat.Audio@orange.fr) - [www.college-nat-audio.fr](http://www.college-nat-audio.fr)

Level »), par rapport à des niveaux de référence (pour chaque fréquence). Du point de vue clinique, une perte auditive à une fréquence audiométrique donnée est observée lorsque l'élévation du seuil absolu d'audition est supérieure à 20 dB HL.

Par convention, chaque degré de perte auditive est défini selon l'intervalle en dB HL présenté ci-après (Goodman, 1965) :

de -10 à 26 dB HL : Normo-entendant

de 27 à 40 dB HL : Pertes légères

de 41 à 55 dB HL : Pertes modérées

de 56 à 70 dB HL : Pertes modérées-sévères

de 71 à 90 dB HL : Pertes sévères

au-delà de 90 dB HL : Pertes profondes (cophoses).

L'élévation des seuils absolus consécutive à une atteinte cochléaire devrait a priori affecter le démasquage de la parole en limitant l'écoute des segments de parole dans les vallées du bruit de fond (*e.g.*, Trine, 1995). Néanmoins, les faibles bénéfices prothétiques en démasquage de la parole que nous discuterons plus loin suggèrent que l'audibilité ne constitue pas un facteur clé de la réduction du démasquage dans les surdités cochléaires (cependant, voir George *et al.*, 2006, pour une étude – la seule à notre connaissance – suggérant l'existence d'un lien clair entre audibilité et démasquage).

### Recrutement de sonie

Une atteinte cochléaire est généralement associée à un phénomène connu sous le nom de *recrutement de sonie* (Fowler, 1936 ; Steinberg and Gardner, 1937), c'est-à-dire une croissance anormalement rapide de l'intensité perçue (la sonie) en fonction de l'intensité physique des sons. Cette croissance rapide est liée à la réduction de l'étendue (la dynamique) des niveaux sonores audibles et confortables. En effet, le seuil d'audibilité du malentendant s'élève alors que le seuil d'inconfort et de douleur reste à un niveau comparable à celui d'un sujet normo-entendant, réduisant ainsi la dynamique audible. En moyenne, la croissance de l'intensité perçue avec l'augmentation de l'intensité physique est d'autant plus

rapide que le seuil absolu du sujet est élevé (Miskolczy-Fodor, 1960 ; Glasberg and Moore, 1989 ; Hellman and Meiselman, 1990 ; 1993). Cependant, lorsque le niveau sonore est suffisamment élevé (supérieur à 90 dB SPL), le niveau d'intensité perçue par un sujet malentendant est similaire à celui du sujet normo-entendant (Moore *et al.*, 1992). Le recrutement de sonie serait associé à une réduction de la compression instantanée (*fast-acting compression*) dans la fonction d'entrée/sortie de la membrane basilaire, et de fait à une véritable expansion ou magnification des fluctuations perceptives d'amplitude dans la cochlée. En effet, une perte de compression d'amplitude devrait conduire à une accélération de la croissance de l'intensité perçue en fonction de l'intensité physique des sons. De plus, la fonction d'entrée/sortie de la membrane basilaire devient presque linéaire pour des niveaux de stimulation supérieurs à 90 dB SPL, expliquant ainsi la similitude entre les fonctions de sonie de sujets malentendants et de sujets normo-entendants à de tels niveaux.

En magnifiant la perception des fluctuations d'amplitude des signaux de parole et du masque, le recrutement de sonie devrait sensiblement affecter le démasquage de la parole des sujets malentendants (en augmentant les phénomènes de masquage de modulation et de MDI entre parole et masque, par exemple). Néanmoins, les rares études ayant estimé le masquage de modulation chez les malentendants suggèrent que ce type de masquage dans le domaine temporel n'est pas affecté par une perte cochléaire (Grose and Hall, 1994 ; Lorenzi *et al.*, 1997b ; Richardson *et al.*, 1998 ; Bacon and Opie, 2002). Par ailleurs, les faibles bénéfices des prothèses à compression en termes de démasquage de la parole suggèrent que le recrutement de sonie ne joue pas un rôle déterminant dans la réduction du démasquage observée dans les surdités cochléaires.

### Perte de sélectivité fréquentielle

Les pertes sensorielles cochléaires se manifestent également par un déficit de la sélectivité fréquentielle (Pick *et al.*, 1977 ; Zwicker and Schorn, 1978 ; Florentine *et al.*, 1980 ; Patterson *et*

*al.*, 1982 ; Glasberg and Moore, 1986). Ce déficit correspond à un élargissement des filtres auditifs, représenté en **figure 1.11**. Sur la partie gauche, l'élargissement du filtre cochléaire accordé sur une fréquence de 1 kHz pour 6 sujets malentendants est notable par rapport à 6 autres sujets normo-entendants ; la partie droite représente la largeur des filtres (en ERB) pour différentes fréquences, en fonction du seuil audiométrique mesuré à la fréquence audio considérée. Le facteur d'élargissement des filtres mesuré est en haut représenté par rapport à la largeur théorique du filtre cochléaire ( $ERB^N$ ), et en bas par rapport à la largeur en ERB des filtres cochléaires mesurés chez des normo-entendants jeunes au même niveau sonore (car la largeur du filtre est également fonction du niveau de présentation). Aucune relation simple n'apparaît entre élévation du seuil absolu d'audition et perte de sélectivité fréquentielle, tant la variabilité entre les sujets est importante. On note toutefois que 1) les filtres ont tendance à s'élargir pour des seuils absolus supérieurs à 30 dB HL, et 2) que pour des pertes modérées-sévères, ce facteur d'élargissement n'excède pas 5. Globalement, une perte modérée est généralement associée à un élargissement des filtres cochléaires par un facteur 2, et une perte sévère par un élargissement des filtres cochléaires par un facteur 4.

Les conséquences perceptives directes de cette perte de sélectivité fréquentielle sont 1) une diminution de l'indépendance perceptive le long de l'axe fréquentiel audio (l'axe tonotopique) et 2) un effet de lissage des indices de place d'excitation dans la cochlée, effet de lissage susceptible d'exacerber les effets de masquage de la parole produits par les bruits ou par tout signal de parole concurrent. Ces effets sont donc susceptibles de dégrader la plupart des mécanismes (listés plus haut) contribuant au démasquage tels que la ségrégation auditive (*e.g.*, Gaudrain, 2008, pages 101-123), le CMR (en diminuant l'indépendance des canaux fréquentiels audio), ou la restauration phonémique (par dégradation de la redondance des signaux). Ces effets ne peuvent à ce jour être compensés par appareillage auditif.

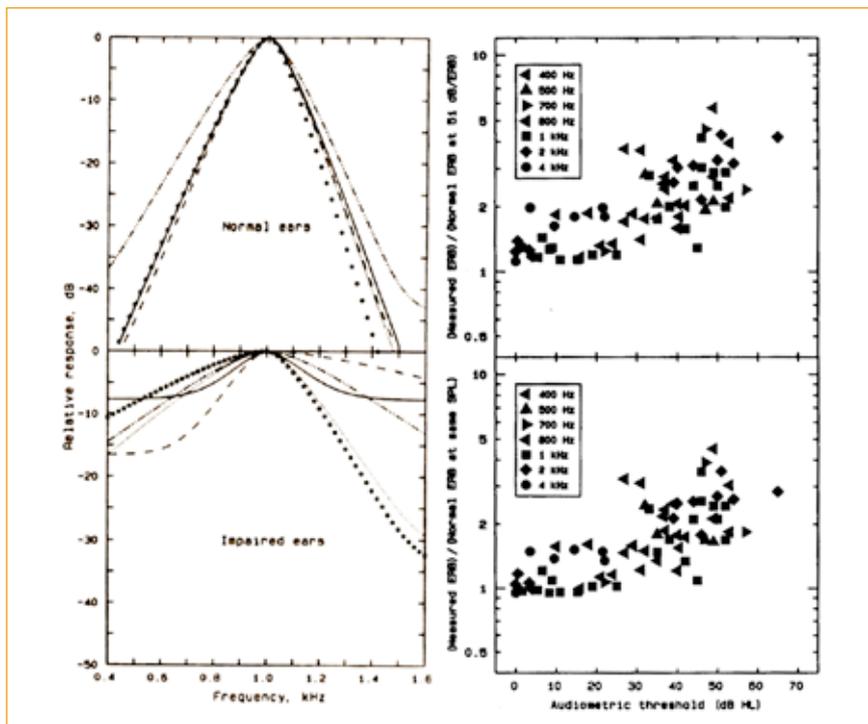


FIGURE 1.11 : A GAUCHE : FORME DES FILTRES AUDITIFS À 1 KHZ POUR 6 OREILLES NORMO-ENTENDANTES (EN HAUT) ET MALENTENDANTES (EN BAS). A DROITE : FACTEUR D'ÉLARGISSEMENT DU FILTRE COCHLÉAIRE PAR RAPPORT À LA LARGEUR THÉORIQUE (ERB™) (EN HAUT), ET PAR RAPPORT À LA LARGEUR MESURÉE CHEZ DES SUJETS NORMO-ENTENDANTS JEUNES AU MÊME NIVEAU SONORE (EN BAS), À DIFFÉRENTES FRÉQUENCES. D'APRÈS MOORE (2007).

### Dégradation de la résolution temporelle et perte de la perception des informations de structure temporelle fine

Enfin, plusieurs études suggèrent un effet des lésions cochléaires sur certaines capacités supraliminaire dans le domaine purement temporel.

Tout d'abord, la résolution ou acuité temporelle, définie comme la limite de la capacité du système auditif à détecter des changements d'enveloppe d'amplitude en fonction du temps, devrait – dans une certaine mesure – être affectée par des lésions cochléaires (Glasberg and Moore, 1992 ; Oxenham and Moore, 1997). En effet, résolution fréquentielle et résolution temporelle ne sont pas indépendantes, et un effet paradoxalement positif de la perte de sélectivité fréquentielle sur la résolution temporelle peut être prédit car la rapidité des fluctuations d'enveloppe en sortie d'un filtre auditif dépend directement de la largeur de ce filtre. En d'autres termes, un élargissement des filtres cochléaires devrait engendrer une amélioration de l'acuité temporelle. Toutefois, une telle amélioration de l'acuité temporelle n'a pas été observée (Moore and Carlyon,

2005). Ces estimations de résolution temporelle ont été réalisées en mesurant la fonction de transfert de modulation temporelle (TMTF) des sujets malentendants. La TMTF estime l'évolution des seuils de détection d'une modulation sinusoïdale d'amplitude en fonction de la fréquence de la modulation, dans le cas de porteuses de type sons purs ou bruits large bande. Les différentes mesures de TMTF suggèrent que la résolution temporelle (estimée par la fréquence de coupure de la TMTF) est préservée chez les sujets malentendants lorsque l'audibilité est contrôlée (Moore, 2007). Toutefois, la sensibilité aux modulations d'amplitude les plus lentes semble améliorée par rapport à la normale, conformément aux effets prévus du recrutement de sonie (une expansion des fluctuations d'amplitude dans la cochlée produite par la perte de la compression cochléaire instantanée).

Enfin, plusieurs études suggèrent qu'une lésion cochléaire pourrait affecter la perception des informations de structure temporelle fine du signal, et dégrader ainsi l'estimation de la fréquence des composantes des signaux de parole réalisée grâce aux propriétés de synchronisation des

impulsions des fibres du nerf auditif (verrouillage de phase évoqué dans une section antérieure) (pour une revue, voir Lorenzi and Moore, 2008 ; et Moore, 2008). Moore et Moore (2003) et Hopkins et Moore (2007) ont étudié les capacités de discrimination de sons complexes harmoniques et inharmoniques aux composantes fréquentielles non résolues (> 5 - 6<sup>ème</sup> harmonique) chez des sujets normo-entendants et des sujets malentendants, en prenant soin d'éliminer toute contribution de sons de combinaison (par masquage) et d'indices de pattern d'excitation (par filtrage). Les bonnes performances de discrimination de ces signaux obtenues chez les sujets normo-entendants suggèrent que ces derniers utilisent des différences de structure temporelle fine. En revanche, les mauvaises performances des sujets malentendants suggèrent fortement que ces derniers ne peuvent pas exploiter ce type d'indice temporel, et se basent uniquement sur des indices d'enveloppe temporelle dans les signaux.

Lorenzi *et al.* (2006a) ont tenté de mesurer l'importance et l'effet de dégradation des indices de structure temporelle fine pour la perception de la parole chez les sujets malentendants. Les stimuli utilisés sont des syllabes sans signification a/C/a, où C est une consonne. Ces logatomes sont divisés en 16 bandes de fréquence audio de largeur égale à 2 ERB, dans lesquelles la structure temporelle fine d'une part et l'enveloppe temporelle d'autre part sont isolées par décomposition de Hilbert. Des signaux véhiculant uniquement les informations de structure temporelle fine ou d'enveloppe temporelle sont alors synthétisés et présentés dans le silence pour identification aux différents groupes de sujets normo-entendants et malentendants jeunes et âgés, après amplification linéaire au demi-gain dans le cas des sujets malentendants. Les résultats (figure 1.12) démontrent que les sujets normo-entendants atteignent de hauts niveaux d'identification des logatomes avec l'enveloppe seule ou la structure temporelle fine seule du signal dans chacune des 16 bandes de fréquence comprises entre 80 et 8000 Hz. En revanche, les sujets malentendants, quel que soit leur âge, semblent

peu capables de tirer profit des informations de structure temporelle fine pour identifier les signaux, alors qu'ils présentent des performances normales avec l'enveloppe seule (conformément aux mesures de perception d'enveloppe (TMTF) discutées plus haut). Cette deuxième étude semble confirmer l'idée selon laquelle une lésion cochléaire engendre un trouble de perception des informations de structure temporelle fine, ou en d'autres termes, une incapacité à encoder et/ou utiliser ces informations de structure temporelle fine véhiculée par les distributions temporelles de potentiels d'action dans les fibres du nerf auditif. Point important, les performances résiduelles en identification de parole de structure temporelle fine des sujets malentendants jeunes s'avèrent fortement corrélées à leurs scores de démasquage mesurés par ailleurs sur les logatomes d'origine en utilisant un bruit stationnaire et un bruit modulé sinusoidalement en amplitude à 8 Hz (100 % de profondeur). Ce constat suggère un rôle important des indices de structure temporelle fine dans le démasquage et le mécanisme d'écoute dans les vallées.

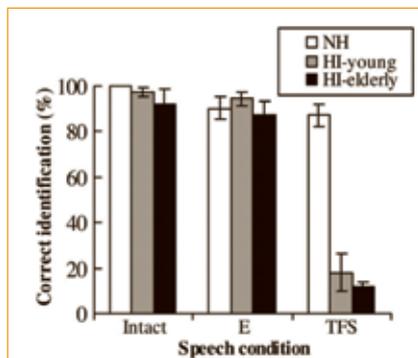


FIGURE 1.12 : POURCENTAGE D'IDENTIFICATION CORRECTE DES LOGATOMES VCV POUR TROIS CONDITIONS : SIGNAL INTACT, ENVELOPPE (E) ET STRUCTURE TEMPORELLE FINE (TFS) POUR CHACUN DES GROUPES DE SUJETS : NORMO-ENTENDANTS (NH), MALENTENDANTS JEUNES (HI-YOUNG) ET MALENTENDANTS ÂGÉS (HI-ELDERLY). D'APRÈS LORENZI ET AL. (2006A).

Une autre étude réalisée par Hopkins *et al.* (2008) démontre plus clairement ce trouble de la perception des indices de structure temporelle fine chez les sujets malentendants. Dans cette étude, les auteurs mesurent le SRT à partir de phrases cibles et d'un masque correspondant à un locuteur interférent chez des sujets normo-entendants et malentendants, en faisant varier la quantité

de structure temporelle fine transmise par bande de fréquence. Pour cela, les auteurs filtrent le mélange (signal de parole et bruit) par un banc de filtres de 32 bandes de largeur 1 ERB, et dégradent l'information de structure temporelle fine dans un nombre croissant de bandes (en la remplaçant par une bande de bruit) en partant des bandes les plus hautes pour aboutir aux bandes les plus basses.

On remarque en **figure 1.13** que, contrairement aux sujets normo-entendants, le SRT ne varie guère chez les sujets malentendants, que l'information de structure temporelle fine transmise soit dégradée ou non au sein d'un nombre croissant de bandes de fréquence. Cette quasi « surdité » aux informations de structure temporelle fine indique qu'en environnement bruité, les sujets malentendants ne parviennent pas à utiliser l'information de structure temporelle fine disponible pour séparer la parole du masque. On notera enfin que conformément aux propriétés de synchronisation des fibres du nerf auditif (se détériorant au dessus de 1-1,5 kHz pour être abolies à 5 kHz), la contribution des indices de structure temporelle fine au SRT se limite aux bandes de fréquence audio inférieures à environ 1,6 kHz chez les sujets normo-entendants. Nous reviendrons plus loin sur cette remarque.

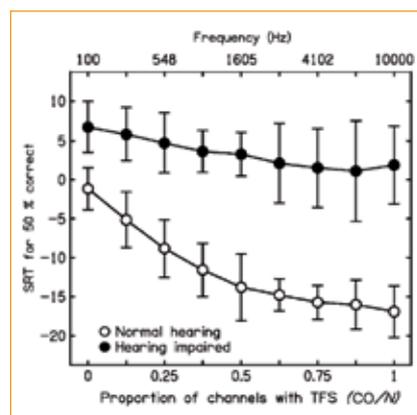


FIGURE 1.13 : SRT MOYEN POUR DES SUJETS NORMO-ENTENDANTS ET MALENTENDANTS EN FONCTION DE LA PROPORTION DE CANAUX (32 AU TOTAL) VÉHICULANT LES INFORMATIONS INITIALES DE STRUCTURE TEMPORELLE FINE. LA FRÉQUENCE CENTRALE DU FILTRE AUDIO AU-DELÀ DE LAQUELLE LA STRUCTURE TEMPORELLE FINE EST DÉGRADÉE EST INDICUÉE SUR L'ABSCISSE DU HAUT. D'APRÈS HOPKINS ET AL. (2008).

## Le démasquage chez les sujets malentendants

Dans le silence, l'intelligibilité de la parole ne diffère guère entre sujets normo-entendants et sujets malentendants jeunes, lorsque l'audibilité est contrôlée par amplification linéaire ou non-linéaire. Cependant, de nombreuses études montrent que **le phénomène de démasquage de la parole est substantiellement réduit chez les personnes atteintes de perte auditive neurosensorielle légère, modérée ou sévère** (*e.g.*, Duquesnoy, 1983 ; Festen and Plomp, 1990 ; Gustafsson and Arlinger, 1994 ; Eisenberg *et al.*, 1995 ; Bacon *et al.*, 1998 ; Peters *et al.*, 1998 ; Dubno *et al.*, 2002 ; Summers and Molis, 2004 ; Lorenzi *et al.*, 2006b).

Pour des sujets malentendants âgés présentant des pertes modérées en haute fréquence et une perte légère en basse fréquence, Duquesnoy (1983) rapporte les mêmes valeurs de SRT pour un bruit stationnaire et pour un locuteur interférent. Le démasquage est donc littéralement aboli ici (malgré une audition quasi-normale en basse fréquence (< 1 kHz) !).

Gustafsson et Arlinger (1994) ont mesuré systématiquement le démasquage de la parole chez un groupe de sujets malentendants âgés présentant des pertes modérées-sévères en haute fréquence en utilisant le paradigme suivant. Le rapport signal-sur-bruit est fixé individuellement afin que chaque sujet atteigne 30 % d'intelligibilité de phrases dans le cas du bruit stationnaire. Ce rapport signal-sur-bruit est ensuite utilisé pour estimer l'intelligibilité des mêmes signaux de parole en présence d'un bruit modulé à l'aide d'une sinusoïde dont la fréquence est variée. Gustafsson et Arlinger remarquent que, comme les sujets normo-entendants, le démasquage mesuré pour ces sujets malentendants présente un pic pour des fréquences de modulation du masque comprises entre 10 et 20 Hz. Toutefois, le niveau de démasquage est réduit de moitié par rapport à un groupe de sujets normo-entendants apparié en âge. Par ailleurs, l'appareillage auditif ne modifie pas les scores de démasquage chez ces sujets malentendants. Ce résultat est confirmé par de nombreuses études

ultérieures examinant plus précisément l'effet de l'appareillage par prothèses auditives à amplification linéaire ou à compression (*e.g.*, Moore *et al.*, 1999 ; Stone *et al.*, 1999).

De la même façon, Lorenzi *et al.* (2006b) ont étudié le démasquage de parole en fonction de la fréquence de modulation du masque (bruit modulé sinusoidalement) pour un groupe de quatre sujets malentendants présentant des pertes plates modérées, en faisant varier la fréquence de modulation de façon systématique entre 2 et 128 Hz. Un faible démasquage n'est observé que pour deux des quatre sujets, avec des valeurs pics pour des fréquences de modulation inférieures à celles relevées pour les sujets normo-entendants (2 - 8 Hz). La **figure 1.14** reprend ces résultats et ceux de Gustafsson et Arlinger (1994).

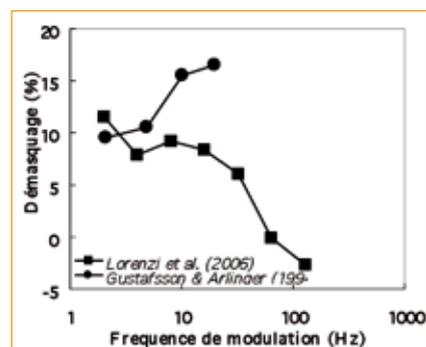


FIGURE 1.14 : DÉMASQUAGE DE LA PAROLE EXPRIMÉ EN TERMES DE DIFFÉRENCE DE POURCENTAGE D'IDENTIFICATION, EN FONCTION DE LA FRÉQUENCE DE MODULATION DU MASQUE. D'APRÈS LORENZI ET AL. (2006B), RÉALISÉ AVEC STIMULI VCV (CARRÉS) ; GUSTAFSSON ET ARLINGER (1994) RÉALISÉ AVEC STIMULI PHRASES (ROND).

L'étude des effets de l'introduction de vallées spectrales et spectro-temporelles dans le bruit masquant menée par Peters *et al.* (1998 ; voir 1.2.c.) a également été conduite sur des sujets malentendants jeunes et âgés, présentant des pertes modérées à modérées-sévères. Pour ces sujets, les auteurs n'observent qu'un démasquage extrêmement réduit dans les situations les plus favorables (-6 dB SRT, par rapport à -19 dB SRT observé chez les sujets normo-entendants), c'est-à-dire dans le cas de vallées temporelles et de vallées spectrales relativement larges (3 ERB<sup>n</sup>).

Pris ensemble, ces résultats démontrent donc que les mécanismes de démasquage de la parole – et plus particulièrement la capacité à extraire les informations de parole disponibles dans les minima temporels et spectraux du bruit masquant – sont fortement affectés par une surdité cochléaire. La perte de l'audibilité semble jouer un rôle mineur dans ce déficit (cependant, voir George *et al.*, 2006 ; cf section 7.2.b, p.194), qui serait plutôt dû à des déficits supraliminaires du traitement temporel et spectral du signal de parole, à savoir un trouble du codage de la structure temporelle fine et une dégradation de la sélectivité fréquentielle (*e.g.*, Bacon *et al.*, 1998 ; Peters *et al.*, 1998 ; Zeng *et al.*, 2005 ; Füllgrabe *et al.*, 2006 ; Lorenzi *et al.*, 2006a). Ces résultats sont conformes aux faibles bénéfices observés dans le cas d'un appareillage par prothèses auditives conventionnelles, ces dernières ne pouvant – à ce jour – compenser le lissage fréquentiel (la dégradation des indices de place d'excitation) et la perte de résolution fréquentielle (le nombre de canaux fréquentiels indépendants) induits par la perte de sélectivité fréquentielle, et restaurer un codage temporel fin normal.

## b. Implant cochléaire

Dans les vingt dernières années, les recherches sur l'implant cochléaire – un dispositif de réhabilitation de la surdité cochléaire relativement récent – ont joué un rôle déterminant et moteur dans l'étude de l'intelligibilité de la parole et le démasquage, et offre une opportunité unique d'estimer la contribution spécifique de certains indices temporels et spectraux au démasquage, car leur transmission au système auditif central est connue et fortement contrainte par le processeur vocal de l'implant dont le principe est détaillé ci-dessous.

### Définition

L'implant cochléaire est un appareil permettant une stimulation électrique de l'oreille interne, afin de restaurer une perception auditive dans le cas de surdités profondes (pour une revue détaillée, voir Clark *et al.*, 1990 ; et Tyler,

1993 ; Zeng *et al.*, 2004b ; Wilson and Dorman, 2008a). Le principe consiste à insérer des électrodes dans la cochlée afin de stimuler directement le nerf auditif par des impulsions électriques. L'implant est composé d'une partie externe correspondant à un micro et à un processeur vocal, et d'une partie interne dans laquelle les électrodes insérées dans la cochlée réalisent la stimulation électrique (voir **figure 1.15**). Ces deux parties communiquent à l'aide d'une antenne aimantée et d'un récepteur placé sous la peau.

Cet appareil effectue une analyse temps/fréquence par bande du signal entrant, chaque électrode correspondant à une bande d'analyse. Les électrodes sont disposées le long de la membrane basilaire et véhiculent l'énergie dans chaque canal en modulant un train d'impulsion de cadence fixe (de même cadence pour tous les canaux). Pour des raisons technologiques liées aux possibles interférences entre électrodes proches, plusieurs stratégies de codage ont été mises en œuvre pour délivrer cette information d'enveloppe (et également Wilson *et al.*, 1991 ; pour le détail des principales stratégies de codage, voir Wilson, 2004).



FIGURE 1.15 : REPRÉSENTATION SCHÉMATIQUE DE LA COUPE DU SYSTÈME AUDITIF PÉRIPHÉRIQUE AVEC UN IMPLANT COCHLÉAIRE. APPARAÎT ICI LE PROCESSEUR VOCAL DE LA PARTIE EXTERNE (1), L'ANTENNE (2), LE RÉCEPTEUR (3), ET LES ÉLECTRODES INSÉRÉES DANS LA COCHLÉE (4).

Ce dispositif est donc basé essentiellement sur le codage fréquentiel tonotopique de la cochlée, mais n'exploite que très partiellement les capacités de codage temporel. Autrement dit, **les implants cochléaires actuels transmettent l'enveloppe spectro-temporelle du signal** (avec une résolution fréquentielle limitée au nombre d'élec-

trodes actives), **mais ne délivrent pas à ce jour la structure temporelle fine.** Conformément au principe de traitement de l'implant, la perception des informations d'enveloppe temporelle est relativement préservée chez un certain nombre de patients, comme cela est suggéré par des études psychoacoustiques mesurant soit les seuils de détection d'enveloppes sinusoïdales (*e.g.*, Shannon, 1992 ; Cazals *et al.*, 1994), soit les performances de discrimination d'enveloppes complexes (*e.g.*, Lorenzi *et al.*, 1997a ; 1998). Au bout de quelques mois post-implantation, les patients implantés cochléaires perçoivent donc correctement la parole dans le silence car, comme nous l'avons vu plus haut, l'enveloppe temporelle dans un nombre limité de bandes est suffisante pour une bonne intelligibilité dans le silence (Shannon *et al.*, 1995 ; Smith *et al.*, 2002). Toutefois, force est de constater que la totalité des patients reste extrêmement gênée dans le bruit.

De tels dispositifs existent depuis les années 70. Les premiers implants analogiques étaient composés d'une seule électrode extracochléaire, ne véhiculant que peu d'information vers l'oreille interne. La perception de la parole avec ce type d'implant était extrêmement limitée, et seuls quelques rares patients exceptionnels pouvaient comprendre quelques mots (Hochmair and Hochmair-Desoyer, 1984). Les implants actuels sont maintenant numériques et multicanaux (comprenant entre 10 et 20 canaux environ), et offrent de bien meilleures performances. Les récentes améliorations ont été faites sur la stimulation électrique elle-même (cadence du train d'impulsion plus rapide, ou variable afin de restituer dans une certaine mesure la structure temporelle fine du signal), et sur la résolution fréquentielle (canaux virtuels ajoutés entre les électrodes pour améliorer la résolution, stratégie dite de « current steering » ; voir Koch *et al.*, 2007 ; Berenstein *et al.*, 2008).

Une meilleure maîtrise de la chirurgie permettant de conserver une certaine audition résiduelle, a rendu possible la conception de nouveaux implants dits « mixtes » combinant stimulation électrique (dans les hautes fréquences) et acoustique (dans les basses fréquences).

En effet, certaines personnes présentant des pertes d'audition importantes dans les hautes fréquences tout en conservant une audition basse fréquence insuffisante mais exploitable (audiogramme passe bas de type « pente de ski ») peuvent bénéficier de ce type d'implant, permettant d'améliorer notablement l'intelligibilité dans les bruits stationnaires et fluctuants fort probablement grâce à la contribution d'indices de place d'excitation et de structure temporelle fine en basse fréquence (*e.g.*, Gantz and Turner, 2003 ; 2004 ; Gantz *et al.*, 2005 ; Chang *et al.*, 2006 ; Cullington and Zeng, 2008 ; Li and Loizou, 2008). Par ailleurs, pour les surdités où la cochlée n'est pas implantable (cochlée ossifiée, nerf auditif sectionné ou atrophié), l'implant du tronc cérébral, délivrant une stimulation électrique directement au noyau cochléaire, fournit une autre solution de réhabilitation dont les bénéficiaires sont toutefois bien inférieurs à ceux obtenus par implants ou prothèses mixtes.

Par la suite, quand il sera question d'implant cochléaire, nous nous limiterons au cas classique de l'implant cochléaire multicanaux, les implants mixtes et du tronc cérébral – plus récents – nécessitant des études dédiées. Par ailleurs, le cas de l'implantation pédiatrique ne sera pas envisagé.

### **L'intelligibilité de la parole dans le bruit sous implant cochléaire**

Les implants cochléaires multicanaux ont évolué au point de permettre une très bonne intelligibilité dans le silence (*e.g.*, Clark *et al.*, 1987 ; Bergeron *et al.*, 1989 ; Tyler *et al.*, 1989), bien qu'une grande variabilité demeure entre les sujets (difficilement contrôlable car elle peut dépendre de facteurs cliniques comme par exemple l'étiologie de la surdité ou la durée de privation sensorielle ; *e.g.*, Cazals *et al.*, 1994). La restauration d'une intelligibilité correcte dans le silence par l'implant cochléaire semble donc être un objectif atteint depuis plus d'une dizaine d'années.

Le nouveau défi de l'implantation cochléaire est maintenant l'amélioration de l'intelligibilité de la parole dans le bruit. En effet, de nombreuses études rapportent d'une part que l'intelligibilité de la parole dans le bruit est très

dégradée, et d'autre part que le démasquage de la parole est presque inexistant (*e.g.*, Nelson *et al.*, 2003 ; Cullington and Zeng, 2008).

Les effets du bruit sur la perception de la parole chez les patients porteurs d'un implant cochléaire ont fait l'objet de nombreuses études, ces dernières portant entre autres sur l'évaluation de différentes stratégies de codage (*e.g.*, Dowell *et al.*, 1987 ; Hochberg *et al.*, 1992 ; Skinner *et al.*, 1994 ; Muller-Deiler *et al.*, 1995 ; Kiefer *et al.*, 1996 ; Fu *et al.*, 1998 ; Zeng and Galvin, 1999). Ainsi, Fu *et al.* (1998) ont mesuré le niveau de bruit stationnaire faisant chuter l'intelligibilité (identification de voyelles et consonnes) de moitié par rapport au score dans le silence (appelé niveau PRT pour « phoneme recognition threshold ») chez un groupe de patients implantés cochléaires et de sujets normo-entendants. Une différence de PRT de plus de 10 dB est observée entre sujets normo-entendants et patients implantés cochléaires.

Le démasquage de la parole chez les patients implantés a été explicitement étudié par Nelson *et al.* (2003). Dans cette étude, ces auteurs mesurent l'intelligibilité de phrases présentées dans un bruit stationnaire, et dans un bruit interrompu périodiquement (grâce à une modulation carrée) et de même énergie chez un groupe de neuf patients implantés cochléaires (porteurs d'implants de marques et stratégies de codage variées). Le démasquage estimé à partir de ces deux conditions (différence de pourcentage de réponse correcte entre le cas fluctuant et le cas stationnaire) est ensuite estimé pour chaque fréquence d'interruption testée et chaque rapport signal-sur-bruit. Les résultats sont présentés en **figure 1.16** : pour des rapports signal-sur-bruit de +8 et +16 dB, une quasi abolition du démasquage est observée chez les patients implantés cochléaires (très faible démasquage observé pour des fréquences d'interruption de 1, 16 et 32 Hz). Ces résultats sont confirmés par une étude similaire réalisée par Fu et Nogaki (2005), estimant toutefois le SRT et non les performances d'identification correcte.

Stickney *et al.* (2004) ont également

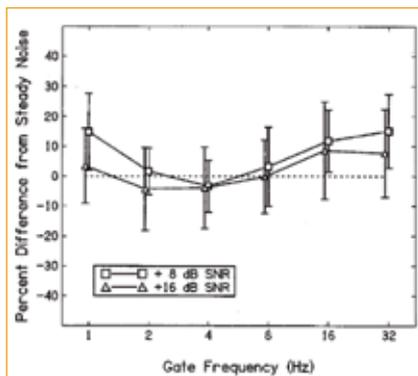


FIGURE 1.16 : DÉMASQUAGE DE LA PAROLE (EXPRIMÉ EN TERMES DE DIFFÉRENCE DE POURCENTAGE D'IDENTIFICATION DE PHRASES OBSERVÉ EN PRÉSENCE DE BRUIT INTERROMPU PÉRIODIQUEMENT ET DE BRUIT STATIONNAIRE) EN FONCTION DE LA FRÉQUENCE DES INTERRUPTIONS DU BRUIT POUR UN GROUPE DE NEUF PATIENTS IMPLANTÉS COCHLÉAIRES, POUR DES RAPPORT SIGNAL-SUR-BRUIT DE +8 ET +16 DB. D'APRÈS NELSON ET AL. (2003).

étudié les mécanismes de masquage de modulation et informationnel chez les patients implantés cochléaires. Dans cette étude, l'intelligibilité de phrases cibles prononcées par un locuteur masculin est estimée en présence de bruit stationnaire ou d'un locuteur interférent de caractéristiques variables (même locuteur que pour la phrase cible, autre locuteur masculin, ou locuteur féminin, répétant toujours la même phrase), et ce pour plusieurs valeurs du rapport signal-sur-bruit. Les résultats sont donnés en **figure 1.17**. Stickney *et al.* remarquent que contrairement à ce qui est observé chez les sujets normo-entendants, l'intelligibilité de la parole en présence d'un locuteur interférent est moins bonne qu'avec le bruit stationnaire. Autrement dit, le démasquage de la parole peut être négatif chez certains patients implantés cochléaires. Les auteurs suggèrent que les sujets implantés intègrent des portions de message venant du masque dans le signal ; ces résultats mettent donc en lumière une contribution des mécanismes de masquage informationnel au démasquage. Par ailleurs, masquer la voix d'homme en utilisant une voix de femme n'améliore pas les performances chez les patients implantés alors que la discriminabilité entre ces deux voix est meilleure (en raison de la plus grande différence de fréquence fondamentale entre les deux voix, supposée conduire à une meilleure ségrégation auditive) ce qui suggère que des mécanismes de

masquage de modulation ont pu également limiter les performances d'identification.

Une autre étude réalisée par le même groupe (Stickney *et al.*, 2007) permet de mieux comprendre la contribution des mécanismes de ségrégation auditive au démasquage chez les sujets normo-entendants et implantés cochléaires. Dans cette étude, l'intelligibilité de phrases est mesurée en présence d'un locuteur interférent pour plusieurs différences de fréquence fondamentale entre locuteurs. Les mauvais résultats obtenus par les patients implantés cochléaires indiquent que les mécanismes impliqués dans la ségrégation auditive de signaux de parole concurrents sont totalement inopérants chez les patients implantés cochléaires.

Nelson et Jin (2004) ont mesuré la robustesse de la perception de la parole aux interruptions, capacité impliquant pour partie des mécanismes de restauration phonémique. Dans cette étude, des phrases interrompues à plusieurs fréquences d'interruption sont présentées pour identification à un groupe de sujets implantés. Les performances d'identification des patients sont au niveau du hasard, révélant une incapacité à inférer les informations acoustico-phonétiques manquantes dans ces conditions expérimentales.

Les mauvaises performances dans

le bruit observées chez les patients implantés peuvent être expliquées par deux déficits majeurs : la mauvaise résolution fréquentielle et la perte des indices structure temporelle fine. En effet, en première approche, la résolution fréquentielle est limitée au nombre d'électrodes insérées et actives dans la cochlée du sujet. Ceci suggère que l'intelligibilité dans le bruit pourrait être substantiellement améliorée en augmentant simplement le nombre d'électrodes : cependant, plusieurs études indiquent que seules huit à dix électrodes sont effectivement utilisées, et que pour des raisons d'interférences entre électrodes, augmenter le nombre d'électrodes au-delà d'une dizaine n'a pas d'effet sur la perception de la parole dans le bruit (Fishman *et al.*, 1997; Friesen *et al.*, 2001).

Par ailleurs, l'implant cochléaire, par principe, ne transmet pas les informations de structure temporelle fine, car la porteuse modulée dans chaque canal est un simple train d'impulsions à cadence fixe. On pourrait penser qu'utiliser une cadence de stimulation variable permettrait aux patients implantés d'avoir accès à certains indices de structure temporelle fine véhiculant par exemple l'information de hauteur fondamentale. Pourtant, la perception de la hauteur sur la base de la cadence de stimulation donne des résultats limités (Zeng, 2002), probablement car la corres-

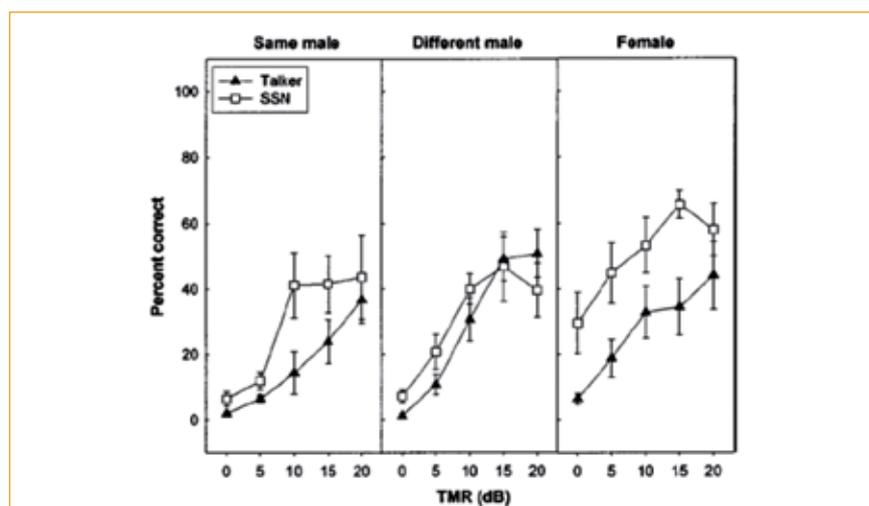


FIGURE 1.17 : INTELLIGIBILITÉ DE PHRASES (LOCUTEUR MASCULIN) MASQUÉES PAR UN AUTRE LOCUTEUR (TALKER, TRIANGLES PLEINS) OU UN BRUIT STATIONNAIRE (SSN, CARRÉS OUVERTS) POUR UN GROUPE DE SUJETS IMPLANTÉS. LE LOCUTEUR INTERFÉRENT EST SOIT LE MÊME QUE CELUI PRONONÇANT LES PHRASES CIBLES (Panneau de gauche), SOIT UN LOCUTEUR MASCULIN DIFFÉRENT (Panneau central), SOIT UN LOCUTEUR FÉMININ. LE BRUIT STATIONNAIRE (SSN) EST DE TYPE SPEECH SHAPED, ET A LE MÊME SPECTRE LONG TERME QUE LE LOCUTEUR INTERFÉRENT DANS CHAQUE CONDITION. D'APRÈS STICKNEY ET AL. (2004).

pondance entre impulsions électriques et potentiels d'action dans les fibres du nerf auditif ne semble pas directe (Dynes and Delgutte, 1992).

En résumé, les études de la perception sous implant cochléaire ont fortement contribué à étayer notre connaissance des mécanismes de démasquage de la parole, en confirmant ou révélant par exemple l'implication des mécanismes de masquage de modulation et des mécanismes de masquage informationnel, ainsi que le rôle des mécanismes de ségrégation auditive (simultanée et/ou séquentielle) et de restauration phonémique. Ces études soulignent l'importance jouée par les indices auditifs de place d'excitation et de structure temporelle fine dans la mise en œuvre efficace de ces différents mécanismes, mais ne permettent pas de séparer leur contribution respective (et différents aspects de leur propre contribution). Les études traitées dans le chapitre suivant, souvent associées à des mesures de la perception sous implant, ont tenté de répondre à cette dernière question.

## CHAPITRE 2 RÔLE DES INDICES AUDITIFS SPECTRAUX ET TEMPORELS DANS LE DÉMASQUAGE

Nous avons vu que les pertes auditives neurosensorielles sont caractérisées par des déficits de sélectivité fréquentielle et de traitement temporel fin, et une réduction, voire une abolition du phénomène de démasquage de la parole. Nous allons maintenant voir, à l'aide de « simulations » sur sujets normo-entendants, comment ces déficits réduisent le démasquage, et ce de façon quantitative. En effet, l'utilisation d'algorithmes de traitement et synthèse du signal de parole testés sur sujets normo-entendants permet de se placer dans un contexte d'étude beaucoup plus contrôlé que les études sur patients dans lesquelles la dégradation des indices spectraux et temporels est bien moins maîtrisée.

Dans cette nouvelle approche, une perte de sélectivité fréquentielle sera simulée grâce à des algorithmes de lissage spectral, et une perte de perception des informations de structure temporelle fine grâce à des vocodeurs.

1

### Effets du lissage spectral

La première étude sur le démasquage de la parole utilisant une technique de lissage spectral est celle menée par ter Keurs *et al.* (1993). Ici, l'algorithme simulant la perte de sélectivité fréquentielle est basé sur une technique d'analyse temps/fréquence appelée « overlap and add » (recouvrement et sommation ; technique permettant d'appliquer la transformée de Fourier inverse de tranches de signal consécutives dans le temps en limitant les effets de bord entre deux tranches, mise au point par Allen, 1977). Le spectre du signal est tout d'abord calculé par transformée de Fourier rapide par tranche de 16,4 ms. Le spectre de puissance est ensuite lissé par filtrage passe bas (convolution sur un échelle logarithmique dans le domaine des fréquences par une fenêtre gaussienne dont la largeur en ERB définit le niveau du lissage) entre 100 et 8000 Hz. Enfin, le signal est reconstitué par transformée de Fourier inverse à partir de ce spectre de puissance lissé et du spectre de phase d'origine (ter Keurs *et al.*, 1992).

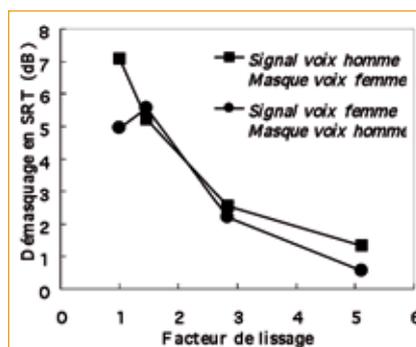


FIGURE 2.1 : DÉMASQUAGE EN SRT (DIFFÉRENCE EN DB SRT ENTRE SRT ESTIMÉ EN PRÉSENCE DE BRUIT FLUCTUANT ET SRT ESTIMÉ EN PRÉSENCE DE BRUIT STATIONNAIRE) EN FONCTION DU FACTEUR DE LISSAGE, D'APRÈS LES DONNÉES DE TER KEURS ET AL. (1993). CE FACTEUR CORRESPOND AU COEFFICIENT D'ÉLARGISSEMENT DU FILTRE D'ANALYSE EN ERB. LA PAROLE CIBLE EST UNE VOIX D'HOMME ET LE MASQUE UNE VOIX DE FEMME (CARRÉS), OU L'INVERSE (RONDS).

Des phrases additionnées à du bruit stationnaire ou à un locuteur interférent sont traitées par ce procédé à plusieurs niveaux de lissage, et présentées pour identification à un groupe de sujets normo-entendants. Le niveau du bruit est maintenu fixe à 65 dB. Les auteurs estiment le rapport signal-sur-bruit conduisant à 50 % de réponses correctes (SRT en dB, vu en section 1.2.a.) en faisant varier le niveau du signal. Les résultats (i.e., le démasquage en dB SRT en fonction du facteur de lissage) sont rapportés en **figure 2.1**. Ici, le facteur de lissage correspond au coefficient d'élargissement en ERB des filtres cochléaires. Ce facteur est calculé à partir de la largeur en octave des fenêtres gaussiennes de lissage (échelle choisie par les auteurs). Les résultats indiquent que pour les faibles valeurs de lissage, le SRT est meilleur lorsque le masque est un locuteur interférent qu'un bruit stationnaire. Autrement dit, un démasquage apparaît clairement pour ces faibles valeurs de lissage. Ce démasquage diminue toutefois à mesure que le lissage augmente, jusqu'à être complètement aboli pour des valeurs de lissage correspondant à une fenêtre gaussienne de plus d'une octave de largeur. En d'autres termes, la perte de sélectivité fréquentielle observée dans les surdités modérées et sévères (conduisant à un élargissement des filtres auditifs par un facteur 2-4) semble capable d'expliquer la forte réduction voire l'abolition du démasquage associée à ces niveaux de surdité.

Un autre algorithme de lissage spectral basé également sur la technique d'« overlap-and-add » a été mis au point par Baer et Moore (1993) et appliqué à l'étude de l'intelligibilité de la parole. Ce procédé est très similaire à celui décrit par ter Keurs *et al.* (1992), mais diffère dans la méthode de lissage du spectre de puissance des signaux. En effet, les auteurs construisent des fonctions de lissage de type « roex » (rounded exponential, fonction suggérée par Patterson *et al.*, 1982), c'est-à-dire des fonctions de lissage dont la forme est celles des filtres auditifs. Cette approche vise ainsi à simuler plus fidèlement les effets perceptifs de l'élargissement des filtres cochléaires chez les sujets malentendants.

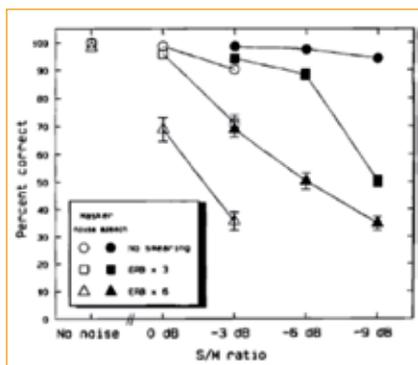


FIGURE 2.2 : INTELLIGIBILITÉ (EXPRIMÉE EN TERMES DE POURCENTAGE DE RÉPONSE CORRECTE) EN FONCTION DU RAPPORT SIGNAL-SUR-BRUIT SANS TRAITEMENT (ROND), AVEC UN LISSAGE SPECTRAL D'UN FACTEUR 3 (CARRÉS), ET D'UN FACTEUR 6 (TRIANGLES). LES SYMBOLES VIDES CORRESPONDENT AUX RÉSULTATS OBTENUS AVEC UN MASQUE STATIONNAIRE (BAER AND MOORE, 1993), ET LES SYMBOLES PLEINS, AVEC UN MASQUE FLUCTUANT (LOCUTEUR INTERFÉRENT). D'APRÈS BAER ET MOORE (1994).

Baer et Moore utilisent ce procédé pour traiter des phrases en présence de bruit stationnaire (Baer and Moore, 1993) ou fluctuant (locuteur interférent ; Baer and Moore, 1994), pour plusieurs rapports signal-sur-bruit et pour deux niveaux de lissage (élargissements d'un facteur 3 et 6). Ces résultats sont rapportés en **figure 2.2** : quel que soit le type de masque et pour chaque rapport signal-sur-bruit, l'intelligibilité se dégrade à mesure que le spectre est lissé. Il est toutefois difficile de comparer les scores d'intelligibilité pour le bruit stationnaire et fluctuant afin d'estimer l'effet du lissage spectral sur le démasquage de la parole, en raison d'un effet plafond (notamment à -3 dB de rapport signal-sur-bruit, où les deux conditions de bruit peuvent être comparées). Les auteurs tentent donc d'estimer le démasquage en termes de rapport signal-sur-bruit conduisant à un niveau d'intelligibilité constant. Ainsi, Baer et Moore remarquent que le rapport signal-sur-bruit conduisant à 70 % de réponses correctes diminue de 3 dB entre les conditions en bruit stationnaire et fluctuant pour un lissage d'un facteur 6, alors que ce rapport diminue de près de 5 dB pour un lissage d'un facteur 3. Le démasquage semble donc moins important pour la plus forte valeur de lissage (facteur 6). Néanmoins et contrairement à l'étude de ter Keurs, le démasquage ne semble pas aboli pour un lissage d'un facteur 6.

Pris ensemble, les résultats de ces différentes études sur les effets du lissage spectral confortent l'idée selon laquelle **les sujets normo-entendants utilisent bel et bien des indices de place d'excitation** pour démasquer la parole. Cette utilisation peut être double : tout d'abord, l'écoute dans les vallées spectrales du masque requiert un certain niveau de résolution fréquentielle ; par ailleurs, des disparités dans les indices de place d'excitation peuvent être utilisées afin de signaler la présence du signal de parole cible dans les vallées du masque. Une perte de sélectivité fréquentielle et le lissage spectral qui en résulte abolissent ces deux composantes du démasquage, expliquant pour une noble part la dégradation du démasquage observée chez les sujets malentendants.

## 2 Etudes sur vocodeurs

### a. Définition d'un vocodeur

Le vocodeur (« vocoder » en anglais) est un dispositif électronique ou informatique de traitement du signal sonore. Son nom vient de la contraction de « *voice coder* » (« codeur de voix »). Un vocodeur analyse les principales composantes spectrales de la voix (ou d'un autre son) et synthétise un son à partir du résultat de cette analyse.



FIGURE 2.3 : DÉMONSTRATION DES PREMIERS VOCODEURS. PHOTOGRAPHIE REPRODUITE DE DUDLEY (1940).

Le premier vocodeur a été imaginé par Dudley *et al.* (1939), ingénieur

aux Laboratoires Bell, afin d'assurer une transmission efficace de la voix dans le cadre du réseau téléphonique américain. Développé initialement sous le nom de « voder » (Voice Operating DEMonstrator), le principe de synthèse est présenté à l'exposition universelle de New York la même année, puis amélioré pour donner lieu à l'actuel vocodeur (Dudley, 1939 ; 1940).

Le principe du vocodeur est de découper le signal de parole en plusieurs bandes de fréquence (ou bandes d'analyse), et de mesurer l'amplitude dans chacune. Seuls ces coefficients sont envoyés, ce afin d'alléger les informations transmises à un réseau téléphonique. Pour reconstruire le message, il faut alors sommer les sorties de plusieurs oscillateurs (autant que de bandes d'analyse, et dont les pulsations correspondent aux fréquences centrales des oscillateurs), pondérées par les coefficients mesurés par le banc d'analyse. Le résultat est intelligible mais, bien entendu, déformé et peu naturel. Cependant, plus le nombre de bandes d'analyse est important, plus la synthèse du modulateur est précise.

Cette technique employée par le vocodeur peut être également appelée « synthèse croisée ». Elle correspond à une décomposition du signal en deux composantes : un signal porteur (oscillateurs) multiplié par un signal modulateur (coefficients de valeur d'énergie dans chaque bande d'analyse). Le vocodeur est utilisé dans plusieurs domaines d'application (télécommunications, musique...) avec différents signaux porteurs, différentes méthodes d'extraction de l'énergie dans une bande, et différents nombres de canaux. Pourtant, le principe reste le même et correspond à une modélisation de l'appareil phonatoire : le signal porteur modélise la soufflerie et le vibrateur, et le signal modulateur le résonateur et l'articulateur. La décomposition du signal selon ce point de vue est appelée modélisation « source/filtre ».

Le vocodeur a été largement utilisé dans des études sur l'intelligibilité de la parole. En effet, cette technique permet de séparer les contributions respectives de l'enveloppe temporelle et de la structure temporelle fine

des signaux, sous l'hypothèse que le système auditif réalise bel et bien une telle décomposition dans le domaine temporel des signaux à bande étroite en sortie des filtres cochléaires. Il faut en effet extraire le signal modulateur en calculant l'enveloppe du signal entrant (module de la transformée de Hilbert, ou rectification et filtrage passe bas), et le signal porteur en extrayant la phase instantanée du signal entrant (argument de la transformée de Hilbert, ou calcul de la modulation de fréquence dans la bande d'analyse).

On peut alors, si l'on choisit un signal porteur sans lien avec la parole d'origine (une bande de bruit ou une sinusoïde à la fréquence centrale du filtre d'analyse) que l'on module ensuite par l'enveloppe, dégrader spécifiquement les indices de structure temporelle fine, et obtenir ainsi un *vocodeur d'enveloppe*. À l'inverse, si l'on ne fournit que la phase instantanée sans lui appliquer de modulation d'amplitude, on obtient un *vocodeur de structure fine*.

Le vocodeur de structure fine a fait l'objet de quelques études (Smith *et al.*, 2002 ; Zeng *et al.*, 2004a ; Gilbert and Lorenzi, 2006 ; Lorenzi *et al.*, 2006a), mais reste peu utilisé car controversé en raison d'une possible reconstruction de l'enveloppe à partir de la structure fine en sortie des filtres cochléaires des auditeurs. Nous n'évoquerons plus le vocodeur de structure fine, et donc par la suite, lorsque nous parlerons d'un vocodeur ou d'un signal vocodé, nous nous limiterons au cas du vocodeur d'enveloppe.

Le vocodeur d'enveloppe extrait donc les modulations d'amplitude lentes du signal de parole, et dégrade les indices de structure temporelle fine. Il effectue de ce fait le même type de transformation que l'implant cochléaire. Ce type de vocodeur constitue par conséquent un outil précieux permettant non seulement de simuler les effets de l'implantation cochléaire, mais également d'étudier les effets de la dégradation des indices temporels sur la perception de la parole. Nous allons d'ailleurs voir que les données de simulation par vocodeur obtenues sur sujets normo-entendants sont très souvent utilisées en complément de données collectées sur patients

implantés. Les vocodeurs permettent également de changer la résolution fréquentielle en modifiant le nombre de canaux, et de lisser les indices de pattern d'excitation en modifiant la pente des filtres d'analyse.

### b. Effet de la résolution fréquentielle

Les effets de la résolution fréquentielle sur le démasquage de la parole ont été étudiés par vocodeur par Qin et Oxenham (2003). Ces auteurs masquent des phrases (locuteur masculin) par différents signaux (bruit stationnaire, locuteurs interférents homme puis femme, et bruit modulé par l'enveloppe du locuteur interférent homme) et les traitent ensuite par un vocodeur (les signaux porteurs étant des bandes de bruit) dont le nombre de canaux est systématiquement manipulé. Qin et Oxenham font donc varier la résolution fréquentielle tout en dégradant la structure temporelle fine, remplacée dans ce cas par du bruit. Le SRT (correspondant à 50 % de réponses correctes) est mesuré sur des sujets normo-entendants pour chaque nombre de canaux du vocodeur et chaque type de masque.

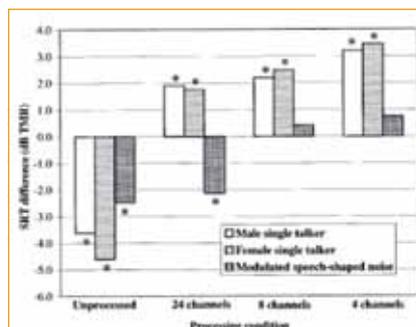


FIGURE 2.4 : DÉMASQUAGE EXPRIMÉ EN TERMES DE DIFFÉRENCE DE SRT ENTRE LES CONDITIONS DE BRUIT STATIONNAIRE ET DE LOCUTEUR INTERFÉRENT HOMME (BARRES OUVERTES), LOCUTEUR INTERFÉRENT FEMME (BARRES GRISES) ET BRUIT MODULÉ (BARRES NOIRES). LE DÉMASQUAGE EST REPRÉSENTÉ POUR DES SIGNAUX INTACTS ET POUR CHAQUE NIVEAU DE RÉOLUTION FRÉQUENTIELLE DU VOCODEUR (4, 8 ET 24 CANAUX). L'ASTÉRISQUE INDIQUE LES VALEURS DE DÉMASQUAGE SIGNIFICATIVEMENT DIFFÉRENTES DE 0 DB. D'APRÈS QIN ET OXENHAM (2003).

Les résultats sont présentés en **figure 2.4** : le démasquage correspond ici à la différence entre les SRT obtenus en présence de masque fluctuant et les SRT obtenus en présence de bruit

stationnaire. Pour la parole intacte (non vocodée), le démasquage est substantiel quel que soit le type de masque fluctuant, conformément aux études antérieures (**voir section 1.2.**). Par contre, lorsque les indices de structure temporelle fine sont dégradés, le démasquage n'est observé qu'en présence de bruit modulé, dans le cas d'un vocodeur de vingt-quatre canaux (pour huit et quatre canaux, aucun démasquage n'est observé avec le bruit modulé). Ceci suggère que les mécanismes d'écoute dans les vallées restent efficaces et permettent de démasquer la parole lorsque la résolution fréquentielle est proche de la normale. Pourtant, en présence d'un locuteur interférent (quel que soit son genre), le démasquage est négatif pour les stimuli traités par le vocodeur, et ce quelle que soit la résolution fréquentielle utilisée. Ceci suggère que les indices de structure temporelle fine sont cruciaux pour la ségrégation de voix concurrentes, car ces indices véhiculent des informations critiques pour l'estimation de la hauteur fondamentale des voix. Par ailleurs, le démasquage négatif observé est compatible avec les résultats obtenus par Stickney *et al.* (2004) sur des sujets implantés : la perte des indices de structure temporelle fine accentue les effets des masquages informationnel et de modulation.

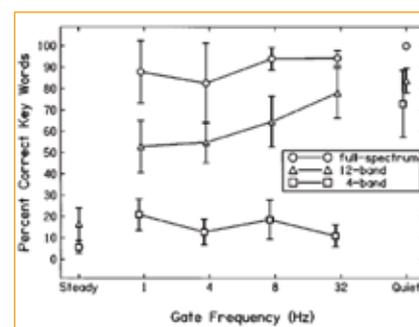


FIGURE 2.5 : INTELLIGIBILITÉ DE PHRASES EN PRÉSENCE DE BRUIT STATIONNAIRE ET DE BRUIT INTERROMPU POUR DES SIGNAUX NON TRAITÉS (CERCLES), ET TRAITÉS PAR UN VOCODEUR DE QUATRE (CARRÉS) ET DOUZE (TRIANGLES) BANDES, EN FONCTION DE LA FRÉQUENCE DES INTERRUPTIONS. D'APRÈS NELSON ET JIN (2004).

Ces résultats sont confirmés par l'étude de Nelson et Jin (2004). Il s'agit ici encore de phrases, traitées par un vocodeur de quatre ou douze bandes remplaçant les informations de structure temporelle fine par des bandes de bruit. Ces stimuli

vocodés sont ensuite présentés dans du bruit stationnaire ou modulé par un signal carré (100 % de profondeur de modulation, temps de cycle de 50 %) à plusieurs fréquences, pour un rapport signal-sur-bruit de -4 dB.

Les résultats obtenus sur sujets normo-entendants (cf. figure 2.5) indiquent que, conformément à l'étude de Qin et Oxenham (2003), un démasquage (différence entre intelligibilité en présence de bruit stationnaire et modulé) n'apparaît que pour une résolution fréquentielle élevée et s'approchant de la normale (ici, douze bandes de fréquence). Toutefois, en l'absence d'indices de structure temporelle fine, le démasquage reste toujours inférieur à celui mesuré pour des stimuli non traités et ce pour les deux niveaux de résolution fréquentielle étudiés.

En résumé, ces résultats suggèrent que **la résolution fréquentielle joue un rôle clé dans le démasquage**, justifiant les efforts menés actuellement par les industriels de l'implant pour augmenter le nombre de canaux fréquentiels de ce dispositif de réhabilitation. Toutefois, **les indices de structure temporelle fine semblent également jouer un rôle critique et distinct**, et ce plus particulièrement lorsque le masque est un signal de parole concurrent.

### c. Effet du lissage spectral

Afin d'étudier plus finement le rôle des indices spectraux (de place d'excitation, pour être plus précis) dans le démasquage de la parole, il est possible de changer les pentes des filtres d'analyse du vocodeur d'enveloppe indépendamment du nombre de canaux ou filtres d'analyse. Le nombre de canaux correspond au facteur de résolution fréquentielle, puisqu'il contrôle l'échantillonnage spectral des informations transmises. Ainsi, plus le nombre de canaux est grand, meilleur est cet échantillonnage. Pourtant, pour un nombre donné de canaux, une modification de la pente des filtres d'analyse affecte le recouvrement spectral entre les canaux (l'indépendance entre canaux) et de fait les informations extraites au sein même d'une bande d'analyse. Cette modification de pente des canaux revient donc à lisser le spectre. Un vocodeur d'en-

veloppe permet donc d'étudier indépendamment les effets produits par un changement de résolution fréquentielle et un changement de lissage spectral (que nous avons discuté en section II.1), et ce en combinaison avec les effets d'une dégradation de structure temporelle fine.

L'effet du lissage spectral sur le démasquage de la parole a été étudié par vocodeur par Fu et Nogaki (2005). Les stimuli sont des phrases présentées dans un bruit stationnaire ou interrompu périodiquement (grâce à une modulation carrée) à plusieurs fréquences. Ces signaux sont ensuite traités par un vocodeur (bandes de bruit comme signal porteur) de 4, 8 ou 16 canaux, pour deux niveaux de lissage spectral. Le traitement du signal diffère quelque peu du lissage spectral que nous venons d'évoquer plus haut, puisque Fu et Nogaki n'ont pas changé la pente des filtres d'analyse, mais celle des filtres de synthèse des bandes de bruits porteuses, à savoir -24 et -6 dB par octave : les auteurs obtiennent ainsi pour des bandes de bruit extraites à -6 dB par octave un spectre plus lissé qu'à -24 dB par octave, et ce pour un même nombre de canaux.

Les résultats sont reproduits en figure 2.6. Conformément aux travaux de Qin et Oxenham (2003) et Nelson et

Jin (2004) discutés précédemment, lorsque le lissage est minimal (pour -24 dB par octave), le démasquage se dégrade lorsque la résolution fréquentielle diminue (environ 12 dB SRT de démasquage pour 16 canaux au maximum, démasquage quasi aboli pour 4 canaux). Les résultats montrent ensuite que le fait de lisser le spectre diminue énormément le démasquage (4 à 5 dB SRT de démasquage pour 16 canaux au maximum, démasquage aboli pour 8 canaux), sans modifier cependant le score dans le bruit stationnaire.

Ces expériences mettent donc en jeu une combinaison des dégradations : en premier lieu, dégradation des indices de structure temporelle fine par vocodeur, et en second lieu, dégradation des indices de place d'excitation par diminution du nombre et de l'indépendance des canaux. Comme pour chaque nouvelle dégradation, le démasquage diminue, ceci suggère que ces indices temporels et spectraux contrôlés par le vocodeur sont déterminants dans le démasquage de la parole. Point important, le masque stationnaire ou fluctuant utilisé dans cette étude est un bruit de type speech-shaped noise, et ne présente pas de vallées spectrales. Le démasquage à l'œuvre dans cette étude est donc de nature purement temporelle (par écoute dans les vallées temporelles

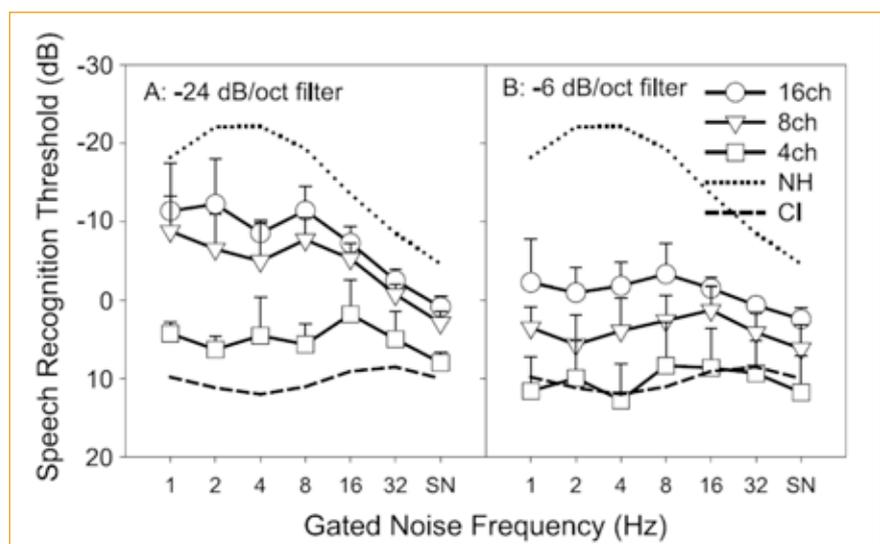


FIGURE 2.6 : SRT EN FONCTION DE LA FRÉQUENCE DES INTERRUPTIONS DU MASQUE POUR UN GROUPE DE SIX SUJETS NORMO-ENTENDANTS (SN POUR BRUIT STATIONNAIRE). LES SIGNAUX SONT TRAITÉS AVEC QUATRE, HUIT ET SEIZE CANAUX. LES BANDES DE BRUIT PORTEUSES SONT EXTRAITES AVEC DES FILTRES RAIDES (A, -24 DB PAR OCTAVE) OU NON (B, -6 DB PAR OCTAVE). LES POINTILLÉS FINS CORRESPONDENT AUX RÉSULTATS OBTENUS EN PAROLE NON TRAITÉE, ET LES POINTILLÉS ÉPAIS, À CEUX OBTENUS SUR SUJETS IMPLANTÉS COCHLÉAIRES. D'APRÈS FU ET NOGAKI (2005).

du bruit). Ces résultats suggèrent donc que des indices de place d'excitation sont utilisés par les sujets normo-entendants afin de déterminer la présence de parole dans les vallées temporelles des bruits masquant.

En résumé, le lissage spectral affecte donc le démasquage de la parole, que ce dernier soit de nature spectrale (écoute dans les vallées spectrales du bruit de fond) ou temporelle (écoute dans les vallées temporelles du bruit de fond). Ce constat justifie les efforts menés actuellement par les industriels de l'implant pour augmenter l'indépendance des canaux fréquentiels de ce dispositif de réhabilitation.

#### d. Effet de la zone fréquentielle traitée

Le vocodeur réalisant un traitement par bande de fréquence, il est alors possible de ne traiter que certaines d'entre elles, et ainsi d'évaluer l'importance des indices d'enveloppe et de structure fine vis-à-vis du démasquage de la parole dans ces différentes zones fréquentielles.

Dans l'étude de Hopkins *et al.* (2008) présentée et discutée précédemment (voir 1.4.a.), un vocodeur est utilisé pour extraire l'enveloppe dans les bandes fréquentielle hautes, sans modifier le signal dans les bandes basses. La limite entre ces zones non modifiées et vocodées est modifiée de façon systématique (signal entièrement traité au minimum, intact au maximum), pour des phrases présentées dans un bruit fluctuant (locuteur interférent). Les résultats (cf. figure 1.12) montrent que l'intelligibilité augmente à mesure que le nombre de bandes vocodées diminue. En d'autres termes, les sujets normo-entendants tirent profit des indices de structure temporelle fine introduits peu à peu dans les basses fréquences. Cependant, les auteurs remarquent que, pour les sujets normo-entendants, les performances s'améliorent clairement jusqu'à 1,6 kHz environ, puis n'augmentent que très faiblement lorsque les indices de structure fine sont introduits au-delà de cette limite. Ceci suggère que les indices de structure temporelle fine n'apportent plus d'information utilisable en plus de l'enveloppe au-delà de 1,6 kHz. Ces

effets observés sont compatibles avec la dégradation de la synchronisation des décharges dans les fibres du nerf auditif au-delà de 1-1,6 kHz (Young and Sachs, 1979 ; Johnson, 1980). Cependant, les études électrophysiologiques révélant un effet de lésions cochléaires sur le verrouillage en phase neural sont rares (*e.g.*, Woolf *et al.*, 1981 ; Miller *et al.*, 1997). Par ailleurs, pour être pleinement probante, cette démonstration devrait être complétée par l'étude des effets d'une dégradation inverse en fréquence.

Dans une autre étude de Hopkins et Moore (2009), le démasquage de la parole est étudié plus explicitement, en utilisant la même méthode. Les résultats observés restent cohérents avec la précédente étude, confirmant ainsi que les capacités de synchronisation des décharges dans les fibres du nerf auditif, et donc la réception des indices de structure temporelle fine, sont cruciales pour le démasquage de la parole.

### 3

#### Comparaison entre lissage spectral et vocodeur

La comparaison des effets d'une dégradation des informations de structure temporelle fine et de la résolution fréquentielle sur le démasquage a été étudiée par Gnansia *et al.* (2009).

Les performances d'identification de consonnes au sein de syllabes sans signification ont été mesurées chez des auditeurs normo-entendants, en présence d'un bruit SSN stationnaire et d'un bruit SSN modulé sinusoidalement en amplitude à 8 Hz. Ici, les performances d'identification ont été mesurées avec un rapport signal-sur-bruit fixé individuellement afin de produire environ 50 % d'intelligibilité en condition de bruit stationnaire. Les signaux (parole et bruit) ont été soit traités par un algorithme de lissage spectral (Baer et Moore, 1993) de façon à dégrader la résolution fréquentielle d'un facteur 2 ou 4, soit traités par un vocodeur (*tone-excited vocoder*) utilisant 32 bandes fréquentielles d'analyse de bande passante ajustable de façon à dégrader à la fois les indices de structure temporelle fine et la résolution fréquentielle.

Cependant, il apparaît clairement que ces algorithmes dégradent à des degrés variables les deux indices, car composantes fréquentielles et temporelles sont intimement liées. Ce point est discuté par Baer et Moore, qui ont estimé qualitativement les effets de leur algorithme de lissage spectral sur la structure temporelle du signal, mais également par Hopkins *et al.* (2008), qui ont discuté la possible transmission d'indices de structure fine par les vocodeurs d'enveloppe (noise vocoders, tone vocoders).

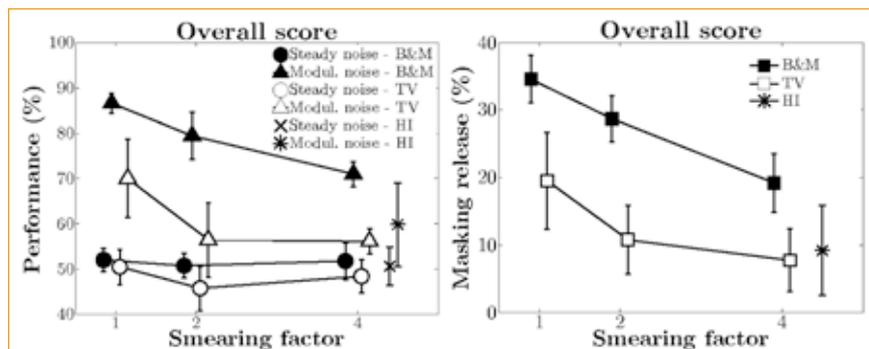


FIGURE 2.7 : A GAUCHE, INTELLIGIBILITÉ DE CONSONNES EN POURCENTAGE DE RÉPONSES CORRECTES, POUR UN GROUPE DE SUJETS NORMO-ENTENDANTS EN PRÉSENCE DE BRUIT STATIONNAIRE (STEADY NOISE, ROND) OU FLUCTUANT (MODUL. NOISE, TRIANGLES), POUR DES SIGNAUX TRAITÉS PAR UN ALGORITHME DE LISSAGE SPECTRAL (B&M, SYMBOLES PLEINS) OU PAR UN VOCODEUR (TV, SYMBOLES OUVERTS). LES PERFORMANCES OBSERVÉES CHEZ DES SUJETS MALENTENDANTS SUR SIGNAUX DE PAROLE NON TRAITÉS (LORENZI ET AL., 2006B) SONT ÉGALEMENT RAPPORTÉS (HI, CROIX AVEC BRUIT STATIONNAIRE, ÉTOILE AVEC BRUIT FLUCTUANT). A DROITE, CES MÊMES RÉSULTATS SONT PRÉSENTÉS EN TERMES DE DÉMASQUAGE DE LA PAROLE (DIFFÉRENCE ENTRE PERFORMANCES OBSERVÉES EN PRÉSENCE DE BRUIT FLUCTUANT ET DE BRUIT STATIONNAIRE), POUR UN GROUPE DE NORMO-ENTENDANTS AVEC DES SIGNAUX DE PAROLE TRAITÉS PAR UN ALGORITHME DE LISSAGE SPECTRAL (B&M, CARRÉS PLEINS), OU UN VOCODEUR (TV, CARRÉS OUVERTS). L'ÉTOILE CORRESPOND AU DÉMASQUAGE OBSERVÉ CHEZ LES MALENTENDANTS. D'APRÈS GNANSIA ET AL. (2009).

Gnansia et coll. ont donc développé un index quantifiant la *fidélité de la transmission des indices d'enveloppe ou de structure temporelle fine*, dans le cas de situations de démasquage (voir également Sheft *et al.*, 2008, pour une autre méthode fondée sur l'extraction de FM au sein de chaque bande). Ceci a permis de constater que la structure temporelle fine est effectivement altérée par l'algorithme de lissage spectral, mais qu'à valeurs de lissage spectral équivalentes (à effets similaires sur les patrons d'excitation spectro-temporels), le vocodeur dégrade plus efficacement la structure temporelle fine.

Les résultats (figure 2.7) ont révélé que les deux algorithmes réduisent le démasquage de la parole, mais l'utilisation du vocodeur affecte plus le démasquage que l'algorithme de lissage spectral à même niveau de dégradation des indices spectraux. Par ailleurs, les auteurs ont observé une correspondance entre le degré de structure temporelle fine transmise et le niveau de démasquage observé. Par ailleurs, l'analyse de la quantité d'information phonétique reçue a révélé une interaction entre effets du lissage fréquentiel et effets de la dégradation de la structure temporelle fine pour le démasquage exprimé en termes de réception du mode d'articulation. Enfin, la comparaison avec des données expérimentales collectées sur sujets malentendants (Lorenzi *et al.*, 2006b) a révélé la nécessité de *combiner lissage fréquentiel et dégradation de la structure temporelle fine* pour rendre pleinement compte des données de démasquage rapportées par ailleurs dans le cas de pertes auditives modérées.

la perception de la parole dans le bruit. Certaines études ont tenté de modéliser ces phénomènes, puis de prédire l'intelligibilité en présence de différents types de bruits stationnaires et fluctuants.

Nous allons nous intéresser à deux modèles, l'un dit *microscopique* et l'autre *macroscopique*. Le premier modèle est microscopique car il tente de modéliser explicitement les processus d'écoute dans les vallées du bruit (Cooke, 2006) ; le second est quant à lui macroscopique car il ne cherche pas à extraire les informations de parole dans les vallées du bruit, mais plutôt à correspondre aux données expérimentales grâce au calcul d'une statistique basée sur l'audibilité du signal de parole par bande de fréquence (Rhebergen and Versfeld, 2005 ; Rhebergen *et al.*, 2006).

### 1 Modèle microscopique du démasquage

Afin de quantifier le phénomène d'écoute dans les vallées du bruit (quantifier le rapport signal-sur-bruit minimum définissant une vallée du bruit pendant laquelle il est possible d'extraire des informations de parole), Cooke (2006) développe un modèle informatique de perception de la parole dans le bruit. L'objectif consiste à modéliser le processus d'écoute dans les vallées du bruit, prédire l'intelligibilité, et faire correspondre les résultats à des données expérimentales (aux résultats sur logatomes VCV de Simpson and Cooke, 2005, voir figure 1.2.). Ainsi, le rapport signal-sur-bruit minimum définissant une vallée de bruit corres-

pond au paramètre du modèle à adapter afin d'ajuster les prédictions du modèle aux données expérimentales.

Le modèle de Cooke (2006) se décompose en trois étapes principales (voir figure 3.1). Les patterns d'excitation spectro-temporels (*Spectrotemporal Excitation Pattern*, STEP ; voir Moore, 2003b) des signaux (parole, bruit, et sommation parole et bruit) sont tout d'abord calculés à l'aide d'un modèle de système auditif périphérique. Cette représentation temps/fréquence contrainte par des facteurs perceptifs permet en effet de déterminer les fluctuations spectro-temporelles de l'enveloppe du signal entrant, et ainsi de repérer des minima temporels ou spectraux de l'enveloppe. Cette détection de minima est ensuite réalisée lors de la seconde étape. Ici, les vallées de bruit sont détectées sur la base d'un critère de rapport signal-sur-bruit minimum choisi. Cette étape de détection des « vallées » (*glimpses*) détermine les régions spectro-temporelles où la parole domine sur le bruit. Ces zones sont enfin traitées par un système de reconnaissance de parole à données manquantes, qui détermine à partir des informations partielles de parole le logatome VCV le plus probable. Ce système de reconnaissance est basé sur des modèles statistiques (modèle de Markov cachés), entraîné au préalable sur de la parole claire (dans le silence). Deux modèles de reconnaissance à données manquantes (*missing data*) sont considérés ici : un premier modèle n'utilisant que l'information partielle mais fiable présente dans les vallées (*glimpses-only model*), et un second modèle utili-

## CHAPITRE 3 MODÈLES MICROSCOPIQUES ET MACROSCOPIQUES DU DÉMASQUAGE

Nous avons présenté et discuté les différents paramètres et mécanismes du démasquage de la parole, ainsi que le rôle clé de certains indices spectraux et temporels, notamment dans le cadre de pertes sensorielles affectant

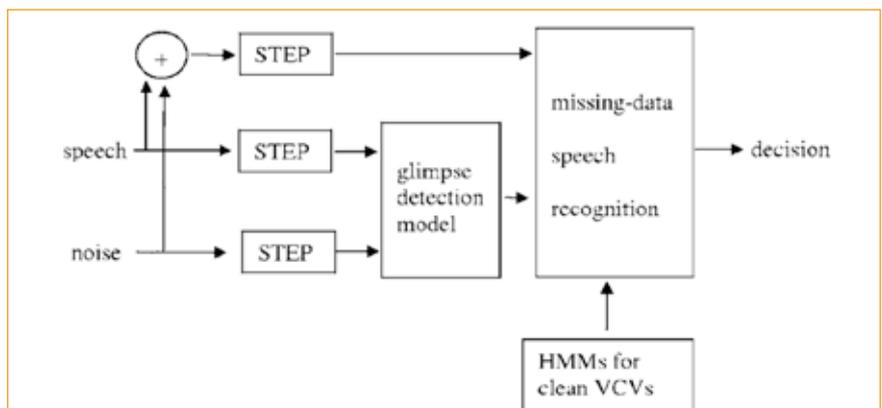


FIGURE 3.1 : ETAPES DE TRAITEMENT DU MODÈLE DE RECONNAISSANCE DE PAROLE DANS LE BRUIT RÉALISANT UNE « ÉCOUTE DANS LES VALLÉES DU BRUIT (GLIMPING) ». D'APRÈS COOKE (2006).

sant également l'information peu fiable présente dans les parties masquées (*glimpses-plus-background model*).

Ce modèle complet est ensuite appliqué à la reconnaissance des stimuli utilisés par Simpson et Cooke (2005), pour plusieurs valeurs de rapport signal-sur-bruit définissant une « vallée ». Pour chacune de ces valeurs, les résultats prédits par le modèle sont comparés aux données comportementales, en calculant pour chaque rapport signal-sur-bruit une distance moyenne et un biais entre les deux résultats. Ces comparaisons révèlent que la distance et le biais sont minimisés pour un rapport signal-sur-bruit de -5 dB (cf. **figure 3.2**) dans le cas du modèle *glimpses-plus-background*, faisant usage des informations partielles de parole recueillies dans les vallées et des informations moins fiables corrompues par le bruit. Toutefois, ces simulations sont également compatibles avec un rapport signal-sur-bruit de -2 dB dans le cas du modèle *glimpses-only*, lorsque l'on ne considère que les fragments de parole contenus dans les vallées. Ces deux valeurs candidates et de fait ces deux modèles de reconnaissance de parole à données manquantes restent à départager. Cooke note toutefois que l'ajustement des données empirique aux simulations favorise le modèle *glimpses-plus-background* (et de fait le rapport signal-sur-bruit seuil de -5 dB).

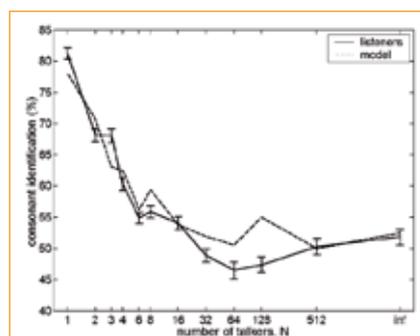


FIGURE 3.2 : IDENTIFICATION DE CONSONNES (VCV, VOYELLE/CONSONNE/VOYELLE) EN FONCTION DU NOMBRE DE LOCUTEURS INTERFÉRENTS POUR UN GROUPE DE SUJETS NORMO-ENTENDANTS (LIGNE CONTINUE) ET POUR LE MODÈLE GLIMPSES-PLUS-BACKGROUND (LIGNE POINTILLÉE). D'APRÈS COOKE (2006).

Cette approche par modélisation permet ainsi de caractériser plus avant le phénomène d'écoute dans les vallées. En effet, nous avons vu précédemment qu'une

vallée de bruit utilisable par le système auditif afin de démasquer la parole est d'une durée comprise entre 16 et 100 ms. Cette étude de modélisation nous indique que sa profondeur minimum est définie par un rapport signal-sur-bruit d'au moins -5 dB ou -2 dB.

Cooke (2006) constate toutefois que, malgré la bonne correspondance entre performances simulées et performances réelles, la réception des traits phonétiques n'est pas correctement prédite par le modèle. En effet, les stimuli sont des logatomes VCV (Voyelle/Consonne/Voyelle, permettant de quantifier la réception des traits phonétiques reçus tels que le voisement, le lieu ou le mode d'articulation), et l'analyse plus précise des résultats révèle de grandes différences entre les données expérimentales et les données simulées. Autrement dit, les logatomes correctement identifiés par les sujets ne correspondent pas toujours à ceux identifiés par le modèle (et inversement). De plus, les mauvaises réponses sont très différentes entre données et prédictions. Tous ces éléments mettent en lumière une limitation importante du modèle.

Il est également important de noter que ce modèle est basé sur l'enveloppe du signal (détection des minima du bruit grâce à des calculs de STEP), et exclut toute contribution des indices de structure temporelle fine dans l'écoute dans les vallées. L'incapacité du modèle à rendre compte des capacités de réception d'information phonétique suggère qu'il pourrait potentiellement être amélioré en intégrant une contribution de ces indices temporels fins.

## 2

### Modèle macroscopique du démasquage, SII étendu

Afin de prédire l'intelligibilité de la parole en présence de bruit stationnaire, French et Steinberg (1947), Fletcher et Galt (1950), et plus tard Kryter (1962a ; b) ont développé une méthode de calcul appelée « index d'articulation » (*Articulation Index, AI*) permettant, à partir de considérations statistiques globales, d'estimer l'intelligibilité de la parole dans un environnement bruité. Dans les années 80, cette

méthode a été examinée de nouveau (Pavlovic, 1984 ; Pavlovic and Studebaker, 1984 ; Kamm *et al.*, 1985 ; Dirks *et al.*, 1986 ; Pavlovic *et al.*, 1986 ; Pavlovic, 1987 ; Studebaker *et al.*, 1987 ; Studebaker *et al.*, 1994), donnant naissance à une méthode affinée ayant fait l'objet d'une standardisation (ANSI, 1997), plus connue sous le nom d'« index d'intelligibilité de la parole » (*Speech Intelligibility Index, SII*). Le SII a finalement été étendu aux bruits fluctuants (Rhebergen and Versfeld, 2005 ; Rhebergen *et al.*, 2006), et permet donc de prédire le démasquage de la parole. C'est ce modèle que nous allons détailler maintenant.

Le SII est basé sur le calcul de la quantité d'information moyenne disponible. Pour cela, le modèle calcule d'abord les spectres de puissance à long terme moyens de la parole et du bruit, divisés en bandes de fréquence. Puis, dans chaque canal, un facteur de correction est appliqué, prenant en compte les effets de masquage énergétique ascendant (*upward spread of masking* ; voir Moore, 2003a) de la parole et du bruit, les seuils d'audibilité, et les effets de distorsion pour les niveaux de présentation élevés. Ensuite, le rapport signal-sur-bruit moyen est calculé dans chaque bande, et ces valeurs sont multipliées par des « fonctions d'importance de bande », qui correspondent à la proportion d'information disponible dans la bande considérée. Ces fonctions dépendent du type de matériel de parole (phrases ou mots) et du niveau de présentation. Enfin, toutes ces valeurs sont additionnées, et conduisent au SII, c'est-à-dire à la quantité d'information de parole disponible.

La méthode de prédiction de l'intelligibilité de la parole dans le bruit fluctuant, appelé « index d'intelligibilité de la parole étendu » (*Extended Speech Intelligibility Index, ESII* ; Rhebergen and Versfeld, 2005 ; Rhebergen *et al.*, 2006), consiste en une extension du SII aux bruits fluctuants.

Le principe de cette méthode a été décrit par Rhebergen et Versfeld (2005), et son architecture est présentée en **figure 3.3**. Il s'agit de calculer le SII sur de courts instants, et de calculer ensuite la moyenne de ces estimations à travers

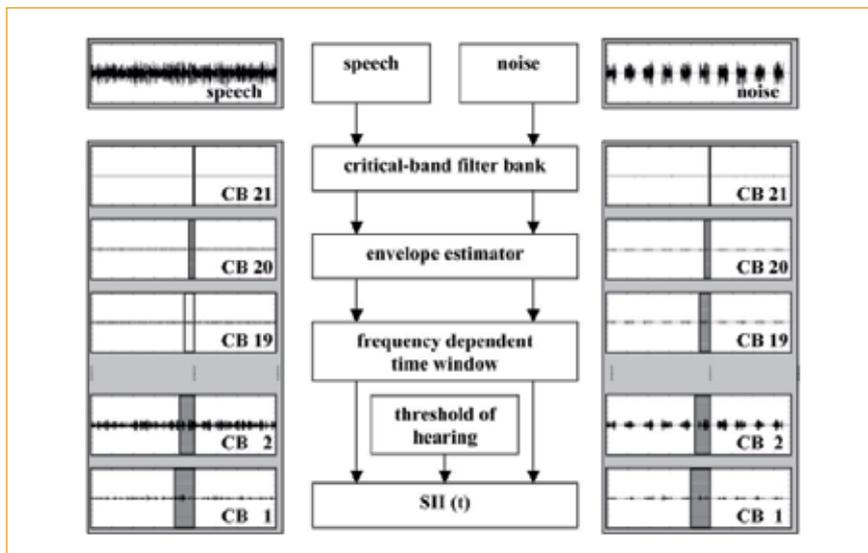


FIGURE 3.3 : REPRÉSENTATION SCHEMATIQUE DE LA CHAÎNE DE TRAITEMENT POUR LE CALCUL DU ESII. LES FORMES D'ONDE DU SIGNAL ET DU BRUIT D'ORIGINE, AINSI QUE DANS CHACUNE DES 21 BANDES DE FRÉQUENCE SONT REPRÉSENTÉES SUR LES CÔTÉS. NOTER EN GRISÉ LES TAILLES DES FENÊTRES DE TRAITEMENT VARIABLES SELON LA FRÉQUENCE CENTRALE DE LA BANDE CONSIDÉRÉE. D'APRÈS RHEBERGEN ET VERSFELD (2005).

le temps. Pour cela, les auteurs modifient quelque peu la méthode de calcul de l'index d'articulation : au lieu de calculer le spectre de puissance et de le découper en plusieurs bandes de fréquence, les signaux (parole et bruit séparément) sont tout d'abord passés dans un banc de filtres (filtres avec des pentes très raides afin de limiter le recouvrement entre les bandes). L'enveloppe de Hilbert est ensuite extraite de chaque signal de bande étroite. Puis, le calcul du SII est réalisé dans des fenêtres temporelles différentes selon la bande de fréquence considérée. Rhebergen et Vesfeld (2005) motivent ce choix par la dépendance fréquentielle de la résolution temporelle du système auditif (Shailer and Moore, 1983 ; 1987). Les fenêtres de traitement sont ainsi plus longues pour les bandes basses que pour les bandes hautes, allant de 35 ms (pour la bande la plus basse, à 150 Hz) à 9,4 ms (pour la bande la plus haute, à 8 kHz). Ces tailles de fenêtres sont définies à partir de mesures psychoacoustiques de seuils de détection d'intervalle de silence (voir Moore, 2003a, Chap 5). Ensuite, le SII est calculé dans chaque intervalle de temps, sans plus de modification par rapport à la méthode originelle, en reprenant les facteurs de correction et les fonctions d'importance de bandes standardisés pour le calcul du SII (ANSI, 1997). Enfin, le ESII est obtenu en moyennant tous les SII calculés au cours du temps.

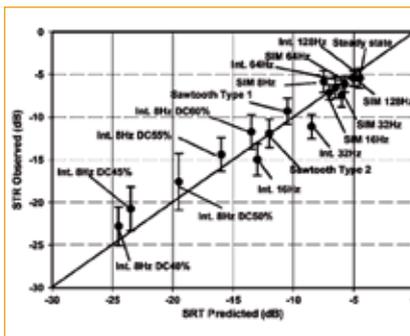


FIGURE 3.4 : SRT OBSERVÉ EN FONCTION DU SRT PRÉDIT (EN DB) PAR LE MODÈLE ESII, POUR PLUSIEURS CONDITIONS DE BRUIT. LA NOMENCLATURE EST LA SUIVANTE : INT. POUR BRUIT INTERROMPU (DIFFÉRENTES FRÉQUENCES, DC EST LE RAPPORT CYCLIQUE, DC = 50% SI NON SPÉCIFIÉ) ; SIM POUR BRUIT MODULÉ PAR UNE SINUSOÏDE (DIFFÉRENTES FRÉQUENCES), SAWTOOTH POUR BRUIT MODULÉ EN DENT DE SCIE (TYPE 1 POUR CROISSANCE EXPONENTIELLE, TYPE 2 POUR DÉCROISSANCE EXPONENTIELLE, À 8 HZ), STEADY STATE POUR LE BRUIT STATIONNAIRE. D'APRÈS RHEBERGEN ET AL. (2006).

Cette méthode ESII est évaluée par Rhebergen *et al.* (2006) sur plusieurs types de bruits : bruit interrompu pour plusieurs fréquences et temps de cycle, bruit modulé par une sinusoïde pour plusieurs fréquences, bruit modulé en dents de scie à 8 Hz à croissance ou décroissance exponentielle, et bruit stationnaire. Pour chaque condition, les auteurs comparent le SRT prédit par le modèle ESII aux données expérimentales collectées sur un groupe de 12 sujets normo-entendants pour

les mêmes stimuli et conditions de bruit (**voir figure 3.4**). Rhebergen *et al.* (2006) constatent que les prédictions sont précises pour une large majorité de conditions de bruit.

Cette modélisation permet donc de bonnes prédictions de l'intelligibilité de la parole dans le bruit stationnaire ou fluctuant. Elle diffère de l'approche microscopique décrite précédemment dans la mesure où aucune hypothèse n'est émise sur les mécanismes de la perception de la parole dans le bruit (notamment aucune hypothèse sur l'écoute dans les vallées). Il s'agit seulement d'une approche statistique par bande de fréquence en fonction du temps. Cependant, cette statistique prend en compte des mesures d'acuité temporelle, laquelle pourrait, comme nous avons pu le voir à plusieurs reprises, jouer un rôle dans le démasquage de la parole.

Ces deux modèles visent à prédire l'intelligibilité de la parole en présence d'un bruit non stationnaire. Cependant, ils impliquent un accès préalable séparé au signal de parole cible et au bruit masquant. Par ailleurs, ces modèles ne réalisent de bonnes prédictions que sur une mesure globale de l'intelligibilité (performances d'identification correcte), mais sont mis en échec lorsqu'il s'agit de prédire le démasquage au niveau plus fin de l'information phonétique reçue (voisement, lieu et mode d'articulation), c'est-à-dire les confusions phonétiques observées sur de véritables sujets. Enfin, ces deux modèles excluent toute information pouvant être extraite de la structure temporelle fine du signal.

## CONCLUSION

### 1 Autres facteurs affectant le démasquage de la parole

Nous avons focalisé notre attention sur l'importance des indices spectraux et temporels du signal de parole dans le démasquage. Nous nous sommes donc limités au déficit de sélectivité fréquentielle et aux possibles troubles de l'utilisation des informations de

verrouillage en phase des décharges neurales dans les fibres du nerf auditif, et avons quelque peu évacué les déficits d'audibilité et la perte de la compression cochléaire instantanée. En effet, le développement des prothèses et implants permet aujourd'hui de corriger convenablement ces déficits (toutefois, sur la vitesse de compression, voir Gilbert *et al.*, 2008) sans pour autant améliorer (significativement) les capacités de démasquage chez les malentendants appareillés. On peut alors raisonnablement considérer que ces déficits d'audibilité et de compression cochléaire ne semblent pas jouer pas un rôle important dans le démasquage. Cette hypothèse a d'ailleurs été confirmée par Eisenberg *et al.* (1995), dans une étude indiquant que l'importante dégradation du démasquage de la parole des sujets malentendants n'est pas observée pour des sujets normo-entendants pour lesquels une perte de l'audibilité est simulée. Cependant, Trine (1995) établit un lien entre audibilité et démasquage, en notant d'une part une corrélation entre la perte d'audibilité et la perte de démasquage chez des sujets malentendants, et en remarquant d'autre part une augmentation du démasquage chez certains sujets malentendants bénéficiant d'une compensation de l'audibilité (par amplification par bande de fréquence) dans certaines conditions. Par la suite, Bacon *et al.* (1998) et Dubno *et al.* (2002), à partir de mesures d'identification de parole dans des masques fluctuants réalisées sur des sujets normo-entendants et malentendants, suggèrent que l'audibilité n'apporte qu'une faible contribution la perte de démasquage de la parole, et que des facteurs supraliminaires sont responsables.

Récemment, une étude de George *et al.* (2006) indique à nouveau un lien possible entre audibilité et démasquage. En effet, des sujets normo-entendants mis en situation de perte d'audibilité (par adjonction d'un bruit masquant stationnaire) montrent une réduction du démasquage. Cette contribution de l'audibilité pourrait de fait diminuer les contributions individuelles et jointes des déficits supraliminaires spectraux et temporels, et modifier substantiellement nos conclusions quant aux effets de pertes cochléaires sur le démas-

quage. Néanmoins, l'usage de bruit masquant pour simuler une élévation de seuils absolus reste tout à fait discutable, entre autres en raison des distorsions de la structure spectro-temporelle du signal (un lissage des fluctuations d'enveloppes spectrales et temporelles, par exemple ; *e.g.*, Steeneken and Houtgast, 1980) induites par le bruit. Par ailleurs, l'adjonction du bruit masquant est susceptible d'augmenter l'amplitude des réponses neurales dans les fibres du nerf auditif, alors qu'une perte d'audibilité devrait engendrer l'effet inverse. De futures expériences semblent donc nécessaires afin de s'assurer de la validité de cette méthode de simulation des troubles d'audibilité.

## 2

### Nouveaux outils d'aide au diagnostic

Le diagnostic et l'évaluation d'une perte auditive reposent encore aujourd'hui sur l'audiogramme tonal. Cependant, les déficits supraliminaires temporels et spectraux et les troubles de démasquage qui en résultent ne sont que très difficilement explicables à partir des seuils audiométriques tonaux. Par ailleurs, des troubles de démasquage importants peuvent être observés chez des patients présentant des seuils normaux ou quasi-normaux sur une plage fréquentielle relativement large (< 2 kHz), et anormaux en hautes fréquences. Ceci suggère que le bilan audiométrique devrait être plus systématiquement complété par une mesure clinique de démasquage de la parole (à savoir, une mesure rapide et simple à mettre en œuvre). Comme nous l'avons vu dans les chapitres précédents, le choix du matériel, la procédure de mesure, et la ou les variables dépendantes sont autant de facteurs clé à prendre en compte dans la réalisation d'un test clinique sensible, pratique et utile.

Un logiciel, nommé « Intellitest », a été développé dans ce but par l'équipe Audition (UMR CNRS 8158) en collaboration avec la société MXM – Neurelec. Il s'agit d'un programme informatique convivial permettant de mettre en œuvre une mesure d'audiométrie vocale dans le bruit à vocation clinique, basée sur des tests d'identification de logatomes VCVCV et CVCVC proches de ceux utilisés lors de nos expériences. L'idée

générale est de pouvoir comparer sujets normo-entendants et malentendants (l'audibilité étant contrôlée directement par le logiciel en utilisant une règle d'amplification proche du demi-gain, mais minimisant les effets de masquage ascendant) dans une tâche d'intelligibilité de parole dans le bruit fournissant un résultat simple en relation avec le niveau de démasquage de la parole.

## 3

### Traitement vocal sous implant

Concernant la correction auditive, nos résultats suggèrent que la transmission de la structure temporelle fine par les appareils auditifs devrait améliorer le démasquage de la parole, et d'une façon plus générale, l'analyse des scènes auditives complexes. Il convient donc dans un premier temps de s'assurer que les prothèses actuelles pouvant a priori transmettre des informations de structure temporelle fine (les informations de phase) le font correctement. Les implants cochléaires, quant à eux, ne le peuvent pas en l'état. Cependant, nous n'apportons pas ici de véritable solution technique à ce problème posé par l'implant. Par ailleurs, la transmission de l'information de structure temporelle fine par le processeur d'une prothèse ou d'un implant est une chose, l'encodage neural normal des informations de structure temporelle fine par verrouillage en phase des décharges des fibres du nerf auditif (la réception de ces informations) en est une autre. En effet, une prothèse auditive peut corriger du mieux possible les déficits des étages de traitement antérieurs à la transduction (restitution de l'audibilité, de la compression cochléaire), c'est au final une cochlée endommagée qui se chargera du codage de ce signal. Peut-être est-il possible de traiter davantage le signal de parole afin d'annuler en amont les effets de ce mauvais codage par la cochlée, mais les quelques essais effectués jusqu'à maintenant n'ont pas été concluants (par exemple sur le rehaussement de l'enveloppe d'amplitude de parole, voir Lorenzi *et al.*, 1999 ; Apoux *et al.*, 2001 ; Apoux *et al.*, 2004). L'implant cochléaire, quant à lui, offre de plus grandes possibilités (Wilson and Dorman, 2008b ; a). En effet, la



- Bregman, A. S.** (1990). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).
- Bregman, A. S., Liao, C., and Levitan, R.** (1990). «Auditory grouping based on fundamental frequency and formant peak frequency,» *Can. J. Psychol.* 44, 400-413.
- Broadbent, D. E., and Ladefoged, P.** (1957). «On the fusion of sounds reaching different sense organs,» *J. Acoust. Soc. Am.* 29, 708-710.
- Bronkhorst, A. W.** (2000). «The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,» *Acustica* 86, 117-128.
- Bronkhorst, A. W., and Plomp, R.** (1992). «Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,» *J. Acoust. Soc. Am.* 92, 3132-3139.
- Brungart, D. S.** (2001). «Informational and energetic masking effects in the perception of two simultaneous talkers,» *J. Acoust. Soc. Am.* 109, 1101-1109.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R.** (2001). «Informational and energetic masking effects in the perception of multiple simultaneous talkers,» *J. Acoust. Soc. Am.* 110, 2527-2538.
- Buss, E., Hall, J. W., 3rd, and Grose, J. H.** (2003). «Effect of amplitude modulation coherence for masked speech signals filtered into narrow bands,» *J. Acoust. Soc. Am.* 113, 462-467.
- Buss, E., Hall, J. W., 3rd, and Grose, J. H.** (2004). «Spectral integration of synchronous and asynchronous cues to consonant identification,» *J. Acoust. Soc. Am.* 115, 2278-2285.
- Carhart, R., Johnson, C., and Goodman, J.** (1975). «Perceptual masking of spondees by combinations of talkers,» *J. Acoust. Soc. Am.* 58, S35-S35.
- Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C.** (1994). «Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants,» *J. Acoust. Soc. Am.* 96, 2048-2054.
- Chang, J. E., Bai, J. Y., and Zeng, F. G.** (2006). «Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise,» *IEEE Trans. Biomed. Eng.* 53, 2598-2601.
- Cherry, E. C.** (1953). «Some experiments on the recognition of speech, with one and with two ears,» *J. Acoust. Soc. Am.* 25, 975-979.
- Clark, G. M., Blamey, P. J., Brown, A. M., Gusby, P. A., Dowell, R. C., Franz, B. K., Pyman, B. C., Shepherd, R. K., Tong, Y. C., and Webb, R. L.** (1987). «The University of Melbourne - Nucleus multi-electrode cochlear implant,» *Adv. Otorhinolaryngol.* 38, V-IX, 1-181.
- Clark, G. M., Tong, Y. C., and Patrick, J. F.** (1990). *Cochlear prostheses* (Churchill Livingstone, Edinburgh).
- Cooke, M.** (2003). «Glimpsing speech,» *J. Phonetics* 31, 579-584.
- Cooke, M.** (2006). «A glimpsing model of speech perception in noise,» *J. Acoust. Soc. Am.* 119, 1562-1573.
- Cooke, M., and Ellis, D. P. W.** (2001). «The auditory organization of speech and other sources in listeners and computational models,» *Speech Communication* 35, 141-177.
- Cullington, H. E., and Zeng, F. G.** (2008). «Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects,» *J. Acoust. Soc. Am.* 123, 450-461.
- Cutler, A., Weber, A., Smits, R., and Cooper, N.** (2004). «Patterns of English phoneme confusions by native and non-native listeners,» *J. Acoust. Soc. Am.* 116, 3668-3678.
- Danhauer, J. L., and Leppler, J. G.** (1979). «Effects of four noise competitors on the California Consonant Test,» *J. Speech. Hear. Disord.* 44, 354-362.
- Darwin, C. J.** (1984). «Perceiving vowels in the presence of another sound: constraints on formant perception,» *J. Acoust. Soc. Am.* 76, 1636-1647.
- Darwin, C. J., and Bethell-Fox, C. E.** (1977). «Pitch continuity and speech source attribution,» *J. Exp. Psychol. Hum. Percept. Perform.* 3, 665-672.
- Dau, T., Kollmeier, B., and Kohlrausch, A.** (1997). «Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,» *J. Acoust. Soc. Am.* 102, 2892-2905.
- de Cheveigné, A.** (1999). «Waveform interactions and the segregation of concurrent vowels,» *J. Acoust. Soc. Am.* 106, 2959-2972.
- Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E.** (1986). «Articulation index predictions of contextually dependent words,» *J. Acoust. Soc. Am.* 80, 82-92.
- Donaldson, G. S., Kreft, H. A., and Litvak, L.** (2005). «Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users (L),» *J. Acoust. Soc. Am.* 118, 623-626.
- Dorman, M. F., Cutting, J. E., and Raphael, L. J.** (1975). «Perception of temporal order in vowel sequences with and without formant transitions,» *J. Exp. Psychol. Hum. Percept. Perform.* 104, 147-153.
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z.** (1998). «The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels,» *J. Acoust. Soc. Am.* 104, 3583-3585.
- Dowell, R. C., Seligman, P. M., Blamey, P. J., and Clark, G. M.** (1987). «Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise,» *Acta Otolaryngol.* 104, 439-446.
- Drullman, R., and Bronkhorst, A. W.** (2000). «Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,» *J. Acoust. Soc. Am.* 107, 2224-2235.
- Drullman, R., Festen, J. M., and Plomp, R.** (1994a). «Effect of reducing slow temporal modulations on speech reception,» *J. Acoust. Soc. Am.* 95, 2670-2680.
- Drullman, R., Festen, J. M., and Plomp, R.** (1994b). «Effect of temporal envelope smearing on speech reception,» *J. Acoust. Soc. Am.* 95, 1053-1064.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B.** (2002). «Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,» *J. Acoust. Soc. Am.* 111, 2897-2907.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B.** (2003). «Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing,» *J. Acoust. Soc. Am.* 113, 2084-2094.
- Dudley, H.** (1939). «The Vocoder,» *Bell Labs. Rec.* 17, 122-126.
- Dudley, H.** (1940). «The carrier nature of speech,» *Bell System Tech. J.* 19, 495-515.
- Dudley, H., Riesz, R. R., and Watkins, S. S. A.** (1939). «A synthetic speaker,» *J. Franklin Inst.* 227, 739-764.
- Duquesnoy, A. J.** (1983). «Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons,» *J. Acoust. Soc. Am.* 74, 739-743.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G.** (2003). «Note on informational masking,» *J. Acoust. Soc. Am.* 113, 2984-2987.
- Dynes, S. B., and Delgutte, B.** (1992). «Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea,» *Hear. Res.* 58, 79-90.

- Eisenberg, L. S., Dirks, D. D., and Bell, T. S.** (1995). «Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing,» *J. Speech Hear. Res.* 38, 222-233.
- Ewert, S. D., and Dau, T.** (2000). «Characterizing frequency selectivity for envelope fluctuations,» *J. Acoust. Soc. Am.* 108, 1181-1196.
- Festen, J. M.** (1987). «Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution,» in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff Publishers, Dordrecht), pp. 461-466.
- Festen, J. M.** (1993). «Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice,» *J. Acoust. Soc. Am.* 94, 1295-1300.
- Festen, J. M., 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.
- Fishman, K. E., Shannon, R. V., and Slattery, W. H.** (1997). «Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,» *J. Speech Lang. Hear. Res.* 40, 1201-1215.
- Fletcher, H., and Galt, R. H.** (1950). «The perception of speech and its relation to telephony,» *J. Acoust. Soc. Am.* 22, 89-151.
- Florentine, M., Buus, S., Scharf, B., and Zwicker, E.** (1980). «Frequency selectivity in normally-hearing and hearing-impaired observers,» *J. Speech Hear. Res.* 23, 646-669.
- Fowler, E. P.** (1936). «A method for the early detection of otosclerosis,» *Arch. Otolaryngol.* 24, 731-741.
- French, N. R., and Steinberg, J. C.** (1947). «Factors governing the intelligibility of speech sounds,» *J. Acoust. Soc. Am.* 19, 90-119.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X.** (2001). «Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants,» *J. Acoust. Soc. Am.* 110, 1150-1163.
- Fu, Q. J., and Nogaki, G.** (2005). «Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing,» *J. Assoc. Res. Otolaryngol.* 6, 19-27.
- Fu, Q. J., Shannon, R. V., and Wang, X.** (1998). «Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing,» *J. Acoust. Soc. Am.* 104, 3586-3596.
- Füllgrabe, C., Berthommier, F., and Lorenzi, C.** (2006). «Masking release for consonant features in temporally fluctuating background noise,» *Hear. Res.* 211, 74-84.
- Gantz, B. J., and Turner, C. W.** (2003). «Combining acoustic and electrical hearing,» *Laryngoscope* 113, 1726-1730.
- Gantz, B. J., and Turner, C. W.** (2004). «Combining acoustic and electrical speech processing: Iowa/Nucleus hybrid implant,» *Acta Otolaryngol.* 124, 344-347.
- Gantz, B. J., Turner, C. W., Gfeller, K. E., and Lowder, M. W.** (2005). «Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing,» *Laryngoscope* 115, 796-802.
- Gaudrain, E.** (2008). «Rôle de la ségrégation séquentielle pour la séparation de voix concurrentes,» Unpublished doctoral dissertation, Claude Bernard University, Lyon.
- Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C.** (2008). «Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation,» *J. Acoust. Soc. Am.* 124, 3076-3087.
- George, E. L., Festen, J. M., and Houtgast, T.** (2006). «Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners,» *J. Acoust. Soc. Am.* 120, 2295-2311.
- Gilbert, G., Akeroyd, M. A., and Gatehouse, S.** (2008). «Discrimination of release time constants in hearing-aid compressors,» *Int. J. Audiol.* 47, 189-198.
- Gilbert, G., and Lorenzi, C.** (2006). «The ability of listeners to use recovered envelope cues from speech fine structure,» *J. Acoust. Soc. Am.* 119, 2438-2444.
- Glasberg, B. R., and Moore, B. C. J.** (1986). «Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,» *J. Acoust. Soc. Am.* 79, 1020-1033.
- Glasberg, B. R., and Moore, B. C. J.** (1989). «Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech,» *Scand. Audiol. Suppl.* 32, 1-25.
- Glasberg, B. R., and Moore, B. C. J.** (1992). «Effects of envelope fluctuations on gap detection,» *Hear. Res.* 64, 81-92.
- Gnansia, D., Jourdes, V., and Lorenzi, C.** (2008). «Effect of masker modulation depth on speech masking release,» *Hear. Res.* 239, 60-68.
- Gnansia, D., Pean, V., Meyer, B., and Lorenzi, C.** (2009). «Effects of spectral smearing and temporal fine structure degradation on speech masking release,» *J. Acoust. Soc. Am.* 125, 4023-4033.
- Goodman, A.** (1965). «Reference zero levels for pure-tones audiometer,» *ASHA* 7, 262-263.
- Grose, J. H., and Hall, J. W., 3rd** (1992). «Comodulation masking release for speech stimuli,» *J. Acoust. Soc. Am.* 91, 1042-1050.
- Grose, J. H., and Hall, J. W., 3rd** (1994). «Modulation Detection Interference (MDI) in listeners with cochlear hearing loss,» *J. Speech Hear. Res.* 37, 680-686.
- Gustafsson, H. A., and Arlinger, S. D.** (1994). «Masking of speech by amplitude-modulated noise,» *J. Acoust. Soc. Am.* 95, 518-529.
- Hall, J. W., 3rd, and Grose, J. H.** (1988). «Comodulation masking release: evidence for multiple cues,» *J. Acoust. Soc. Am.* 84, 1669-1675.
- Hall, J. W., 3rd, Grose, J. H., and Haggard, M. P.** (1988). «Comodulation masking release for multicomponent signals,» *J. Acoust. Soc. Am.* 83, 677-686.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A.** (1984). «Detection in noise by spectro-temporal pattern analysis,» *J. Acoust. Soc. Am.* 76, 50-56.
- Hellman, R. P., and Meiselman, C. H.** (1990). «Loudness relations for individuals and groups in normal and impaired hearing,» *J. Acoust. Soc. Am.* 88, 2596-2606.
- Hellman, R. P., and Meiselman, C. H.** (1993). «Rate of loudness growth for pure tones in normal and impaired hearing,» *J. Acoust. Soc. Am.* 93, 966-975.
- Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S.** (1992). «Effects of noise and noise suppression on speech perception by cochlear implant users,» *Ear. Hear.* 13, 263-271.
- Hochmair, E. S., and Hochmair-Desoyer, I. J.** (1984) in *Cochlear Implants*, edited by R. A. Shindler, and M. M. Merzenich (Raven, New York), pp. 101-110.
- Hopkins, K., and Moore, B. C. J.** (2007). «Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information,» *J. Acoust. Soc. Am.* 122, 1055-1068.
- Hopkins, K., and Moore, B. C. J.** (2009). «The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise,» *J. Acoust. Soc. Am.* 125, 442-446.

- Hopkins, K., Moore, B. C. J., and Stone, M. A.** (2008). «Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech,» *J. Acoust. Soc. Am.* 123, 1140-1153.
- Horii, Y., House, A. S., and Hughes, G. W.** (1970). «Development and evaluation of a noise with speech-envelope characteristics,» *J. Acoust. Soc. Am.* 47, 75.
- Houtgast, T.** (1989). «Frequency selectivity in amplitude-modulation detection,» *J. Acoust. Soc. Am.* 85, 1676-1680.
- Howard-Jones, P. A., and Rosen, S.** (1993a). «The Perception of Speech in Fluctuating Noise,» *Acustica* 78, 258-272.
- Howard-Jones, P. A., and Rosen, S.** (1993b). «Unmodulated glimpsing in «checkerboard» noise,» *J. Acoust. Soc. Am.* 93, 2915-2922.
- Iyer, N., Brungart, D. S., and Simpson, B. D.** (2007). «Effects of periodic masker interruption on the intelligibility of interrupted speech,» *J. Acoust. Soc. Am.* 122, 1693.
- Johnson, D. H.** (1980). «The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,» *J. Acoust. Soc. Am.* 68, 1115-1122.
- Kamm, C. A., Dirks, D. D., and Bell, T. S.** (1985). «Speech recognition and the Articulation Index for normal and hearing-impaired listeners,» *J. Acoust. Soc. Am.* 77, 281-288.
- Kiefer, J., Muller, J., Pfennigdorff, T., Schon, F., Helms, J., von Ilberg, C., Baumgartner, W., Gstottner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., and Baur, S.** (1996). «Speech understanding in quiet and in noise with the CIS speech coding strategy (MED-EL Combi-40) compared to the multipeak and spectral peak strategies (nucleus),» *ORL J. Otorhinolaryngol. Relat. Spec.* 58, 127-135.
- Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L.** (2007). «Using current steering to increase spectral resolution in CII and HiRes 90K users,» *Ear. Hear.* 28, 38S-41S.
- Kryter, K. D.** (1962a). «Methods for the calculation and use of the Articulation Index,» *J. Acoust. Soc. Am.* 34, 1689-1697.
- Kryter, K. D.** (1962b). «Validation of the Articulation Index,» *J. Acoust. Soc. Am.* 34, 1698-1702.
- Kwon, B. J.** (2002). «Comodulation masking release in consonant recognition,» *J. Acoust. Soc. Am.* 112, 634-641.
- Kwon, B. J., and Turner, C. W.** (2001). «Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference ?,» *J. Acoust. Soc. Am.* 110, 1130-1140.
- Lackner, J. R., and Goldstein, L. M.** (1974). «Primary auditory stream segregation of repeated consonant-vowel sequences,» *J. Acoust. Soc. Am.* 56, 1651-1652.
- Li, N., and Loizou, P. C.** (2008). «A glimpsing account for the benefit of simulated combined acoustic and electric hearing,» *J. Acoust. Soc. Am.* 123, 2287-2294.
- Licklider, J. C. R., and Guttman, N.** (1957). «Masking of speech by line-spectrum interference,» *J. Acoust. Soc. Am.* 29, 287-296.
- Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N.** (1999). «Effects of envelope expansion on speech recognition,» *Hear. Res.* 136, 131-138.
- Lorenzi, C., Gallego, S., and Patterson, R. D.** (1997a). «Discrimination of temporal asymmetry in cochlear implantees,» *J. Acoust. Soc. Am.* 102, 482-485.
- Lorenzi, C., Gallego, S., and Patterson, R. D.** (1998). «Amplitude compression in cochlear implants artificially restricts the perception of temporal asymmetry,» *Br. J. Audiol.* 32, 367-374.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J.** (2006a). «Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,» *Proc. Natl. Acad. Sci. U S A* 103, 18866-18869.
- Lorenzi, C., Husson, M., Ardoint, M., and Debrulle, X.** (2006b). «Speech masking release in listeners with flat hearing loss: effects of masker fluctuation rate on identification scores and phonetic feature reception,» *Int. J. Audiol.* 45, 487-495.
- Lorenzi, C., Michey, C., Berthommier, F., and Portalier, S.** (1997b). «Modulation masking in listeners with sensorineural hearing loss,» *J. Speech Lang. Hear. Res.* 40, 200-207.
- Lorenzi, C., and Moore, B. C. J.** (2008). «Role of temporal envelope and fine structure cues in speech perception: A review,» in *International Symposium on Auditory and Audiological Research*, edited by T. Dau (Helsingor, Denmark), pp. 263-273.
- Mackersie, C. L., Prida, T. L., and Stiles, D.** (2001). «The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss,» *J. Speech Lang. Hear. Res.* 44, 19-28.
- McDermott, H. J., and McKay, C. M.** (1994). «Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea,» *J. Acoust. Soc. Am.* 96, 155-162.
- Miller, G. A.** (1947). «The masking of speech,» *Psychol. Bull.* 44, 105-129.
- Miller, G. A., and Heise, G. A.** (1950). «The trill threshold,» *J. Acoust. Soc. Am.* 22, 637-638.
- Miller, G. A., and Licklider, J. C. A.** (1950). «The intelligibility of interrupted speech,» *J. Acoust. Soc. Am.* 22, 167-173.
- Miller, G. A., and Nicely, P. E.** (1955). «An analysis of perceptual confusions among some english consonants,» *J. Acoust. Soc. Am.* 27, 338-352.
- Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D.** (1997). «Effects of acoustic trauma on the representation of the vowel «eh» in cat auditory nerve fibers,» *J. Acoust. Soc. Am.* 101, 3602-3616.
- Millman, R. E., Lorenzi, C., Apoux, F., Fullgrabe, C., Green, G. G., and Bacon, S. P.** (2002). «Effect of duration on amplitude-modulation masking,» *J. Acoust. Soc. Am.* 111, 2551-2554.
- Miskolczy-Fodor, F.** (1960). «Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function,» *J. Acoust. Soc. Am.* 32, 486-492.
- Moore, B. C. J.** (2003a). *An Introduction to the Psychology of Hearing*, Fifth Edition (Academic Press).
- Moore, B. C. J.** (2003b). «Temporal integration and context effects in hearing,» *J. Phonetics* 31, 564-574.
- Moore, B. C. J.** (2007). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (Wiley, Chichester).
- Moore, B. C. J.** (2008). «The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people,» *J. Assoc. Res. Otolaryngol.* 9, 399-406.
- Moore, B. C. J., and Carlyon, R. P.** (2005). «Perception of pitch by people with cochlear hearing loss and cochlear implant users,» in *Pitch: Neural coding and perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer).
- Moore, B. C. J., Johnson, J. S., Clark, T. M., and Pluinage, V.** (1992). «Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss,» *Ear. Hear.* 13, 349-370.

- Moore, B. C. J., and Moore, G. A.** (2003). «Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects,» *Hear. Res.* 182, 153-163.
- Moore, B. C. J., Peters, R. W., and Stone, M. A.** (1999). «Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips,» *J. Acoust. Soc. Am.* 105, 400-411.
- Muller-Deiler, J., Schmidt, B. J., and Rudert, H.** (1995). «Effects of noise on speech discrimination in cochlear implant patients,» *Ann. Otol. Rhinol. Laryngol. Suppl.* 166, 303-306.
- Nelson, P. B., and Jin, S. H.** (2004). «Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners,» *J. Acoust. Soc. Am.* 115, 2286-2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A.** (2003). «Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners,» *J. Acoust. Soc. Am.* 113, 961-968.
- Nie, K., Stickney, G., and Zeng, F. G.** (2005). «Encoding frequency modulation to improve cochlear implant performance in noise,» *IEEE Trans. Biomed. Eng.* 52, 64-73.
- Nooteboom, S. G., Brokx, J. P. L., and de Rooij, J. J.** (1978). «Contribution of prosody to speech perception,» in *Studies in the perception of language*, edited by W. J. M. Levelt, and G. B. F. d'Arcais (Wiley and Sons, New York, USA).
- Oxenham, A. J., and Moore, B. C. J.** (1997). «Modeling the effects of peripheral nonlinearity in normal and impaired hearing,» in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ).
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R.** (1982). «The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold,» *J. Acoust. Soc. Am.* 72, 1788-1803.
- Pavlovic, C. V.** (1984). «Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment,» *J. Acoust. Soc. Am.* 75, 1253-1258.
- Pavlovic, C. V.** (1987). «Derivation of primary parameters and procedures for use in speech intelligibility predictions,» *J. Acoust. Soc. Am.* 82, 413-422.
- Pavlovic, C. V., and Studebaker, G. A.** (1984). «An evaluation of some assumptions underlying the articulation index,» *J. Acoust. Soc. Am.* 75, 1606-1612.
- Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L.** (1986). «An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,» *J. Acoust. Soc. Am.* 80, 50-57.
- Peters, R. W., Moore, B. C. J., and Baer, T.** (1998). «Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,» *J. Acoust. Soc. Am.* 103, 577-587.
- Pick, G., Evans, E. F., and Wilson, J. P.** (1977). «Frequency resolution in patients with hearing loss of cochlear origin,» in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans, and J. P. Wilson (Academic, London).
- Plomp, R., and Mimpen, A. M.** (1979). «Improving the reliability of testing the speech reception threshold for sentences,» *Audiology* 18, 43-52.
- Powers, G. L., and Speaks, C.** (1973). «Intelligibility of temporally interrupted speech,» *J. Acoust. Soc. Am.* 54, 661-667.
- Powers, G. L., and Wilcox, J. C.** (1977). «Intelligibility of temporally interrupted speech with and without intervening noise,» *J. Acoust. Soc. Am.* 61, 195-199.
- Pressnitzer, D., and Hupé, J. M.** (2006). «Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization,» *Curr. Biol.* 16, 1351-1357.
- Qin, M. K., and Oxenham, A. J.** (2003). «Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,» *J. Acoust. Soc. Am.* 114, 446-454.
- Rhebergen, K. S., and Versfeld, N. J.** (2005). «A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners,» *J. Acoust. Soc. Am.* 117, 2181-2192.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A.** (2005). «Release from informational masking by time reversal of native and non-native interfering speech,» *J. Acoust. Soc. Am.* 118, 1274-1277.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A.** (2006). «Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise,» *J. Acoust. Soc. Am.* 120, 3988-3997.
- Richardson, L. M., Busby, P. A., and Clark, G. M.** (1998). «Modulation detection interference in cochlear implant subjects,» *J. Acoust. Soc. Am.* 104, 442-452.
- Riss, D., Arnoldner, C., Baumgartner, W. D., Kaider, A., and Hamzavi, J. S.** (2008). «A new fine structure speech coding strategy: speech perception at a reduced number of channels,» *Otol Neurotol* 29, 784-788.
- Rosen, S.** (1992). «Temporal information in speech: acoustic, auditory and linguistic aspects,» *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 336, 367-373.
- Scheffers, M. T. M.** (1983). «Sifting vowels: Auditory pitch analysis and sound segregation,» Unpublished doctoral dissertation, University of Groningen, The Netherlands.
- Shailer, M. J., and Moore, B. C. J.** (1983). «Gap detection as a function of frequency, bandwidth, and level,» *J. Acoust. Soc. Am.* 74, 467-473.
- Shailer, M. J., and Moore, B. C. J.** (1987). «Gap detection and the auditory filter: phase effects using sinusoidal stimuli,» *J. Acoust. Soc. Am.* 81, 1110-1117.
- Shannon, R. V.** (1992). «Temporal modulation transfer functions in patients with cochlear implants,» *J. Acoust. Soc. Am.* 91, 2156-2164.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M.** (1995). «Speech recognition with primarily temporal cues,» *Science* 270, 303-304.
- Sheft, S., Ardoint, M., and Lorenzi, C.** (2008). «Speech identification based on temporal fine structure cues,» *J. Acoust. Soc. Am.* 124, 562-575.
- Simpson, S. A., and Cooke, M.** (2005). «Consonant identification in N-talker babble is a nonmonotonic function of N,» *J. Acoust. Soc. Am.* 118, 2775-2778.
- Singh, P. G.** (1987). «Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre?,» *J. Acoust. Soc. Am.* 82, 886-899.
- Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., and et al.** (1994). «Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,» *Am. J. Otol.* 15 Suppl 2, 15-27.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J.** (2002). «Chimaeric sounds reveal dichotomies in auditory perception,» *Nature* 416, 87-90.
- Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R.** (2002). «Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity,» *J. Acoust. Soc. Am.* 112, 720-727.
- Steeneken, H. J., and Houtgast, T.** (1980). «A physical method for measuring speech-transmission quality,» *J. Acoust. Soc. Am.* 67, 318-326.
- Steinberg, J. C., and Gardner, M. B.** (1937).

«The dependence of hearing impairment on sound intensity,» *J. Acoust. Soc. Am.* 9, 11-23.

**Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G.** (2007). «Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences,» *J. Acoust. Soc. Am.* 122, 1069-1078.

**Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P.** (2004). «Cochlear implant speech recognition with speech maskers,» *J. Acoust. Soc. Am.* 116, 1081-1091.

**Stohl, J. S., Throckmorton, C. S., and Collins, L. M.** (2008). «Assessing the pitch structure associated with multiple rates and places for cochlear implant users,» *J. Acoust. Soc. Am.* 123, 1043-1053.

**Stone, M. A., Moore, B. C. J., Alcantara, J. I., and Glasberg, B. R.** (1999). «Comparison of different forms of compression using wearable digital hearing aids,» *J. Acoust. Soc. Am.* 106, 3603-3619.

**Strickland, E. A., and Viemeister, N. F.** (1996). «Cues for discrimination of envelopes,» *J. Acoust. Soc. Am.* 99, 3638-3646.

**Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L.** (1987). «A frequency importance function for continuous discourse,» *J. Acoust. Soc. Am.* 81, 1130-1138.

**Studebaker, G. A., Taylor, R., and Sherbecoe, R. L.** (1994). «The effect of noise spectrum on speech recognition performance-intensity functions,» *J. Speech Hear. Res.* 37, 439-448.

**Summers, V., and Leek, M. R.** (1998). «Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners,» *Hear. Res.* 118, 139-150.

**Summers, V., and Molis, M. R.** (2004). «Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level,» *J. Speech Lang. Hear. Res.* 47, 245-256.

**Takahashi, G. A., and Bacon, S. P.** (1992). «Modulation detection, modulation masking, and speech understanding in noise in the elderly,» *J. Speech Hear. Res.* 35, 1410-1421.

**ter Keurs, M., Festen, J. M., and Plomp, R.** (1992). «Effect of spectral envelope smearing on speech reception. I,» *J. Acoust. Soc. Am.* 91, 2872-2880.

**ter Keurs, M., Festen, J. M., and Plomp, R.** (1993). «Effect of spectral envelope smearing on speech reception. II,» *J. Acoust. Soc. Am.* 93, 1547-1552.

**Terhardt, E.** (1972a). «Zur Tonhöhenwahrnehmung

von Klängen. I. Psychoakustische Grundlagen,» *Acustica* 26, 173-186.

**Terhardt, E.** (1972b). «Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionenschema,» *Acustica* 26, 187-199.

**Thomas, I. B., Hill, P. B., Carroll, F. S., and Garcia, B.** (1970). «Temporal order in the perception of vowels,» *J. Acoust. Soc. Am.* 48, 1010-1013.

**Throckmorton, C. S., Selin Kucukoglu, M., Remus, J. J., and Collins, L. M.** (2006). «Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: implications for cochlear implants,» *Hear. Res.* 218, 30-42.

**Trine, T. D.** (1995). «Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth,» Unpublished doctoral dissertation, University of Minnesota, Twin Cities, Twin Cities.

**Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A.** (2004). «Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing,» *J. Acoust. Soc. Am.* 115, 1729-1735.

**Tyler, R. S.** (1993). Cochlear implants: Audiological foundations. (Singular Publishing Group, San Diego, USA).

**Tyler, R. S., Tye-Murray, N., and Otto, S. R.** (1989). «The recognition of vowels differing by a single formant by cochlear-implant subjects,» *J. Acoust. Soc. Am.* 86, 2107-2112.

**van Noorden, L. P. A. S.** (1975). «Temporal coherence in the perception of tone sequences,» Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.

**Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P.** (1987). «Speech waveform envelope cues for consonant recognition,» *J. Acoust. Soc. Am.* 82, 1152-1161.

**Versfeld, N. J., and Dreschler, W. A.** (2002). «The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners,» *J. Acoust. Soc. Am.* 111, 401-408.

**Warren, R. M.** (1970). «Perceptual restoration of missing speech sounds,» *Science* 167, 392-393.

**Wilson, B. S.** (2004). «Engineering design of cochlear implants,» in Cochlear implants, auditory prostheses and electric hearing, edited by F. G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York).

**Wilson, B. S., and Dorman, M. F.** (2008a). «Cochlear implants: a remarkable past and a brilliant future,» *Hear. Res.* 242, 3-21.

**Wilson, B. S., and Dorman, M. F.** (2008b). «Cochlear implants: current designs and future possibilities,» *J Rehabil Res Dev* 45, 695-730.

**Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M.** (1991). «Better speech recognition with cochlear implants,» *Nature* 352, 236-238.

**Woolf, N. K., Ryan, A. F., and Bone, R. C.** (1981). «Neural phase-locking properties in the absence of cochlear outer hair cells,» *Hear. Res.* 4, 335-346.

**Yost, W. A., and Sheft, S.** (1989). «Across-critical-band processing of amplitude-modulated tones,» *J. Acoust. Soc. Am.* 85, 848-857.

**Yost, W. A., and Sheft, S.** (1994). «Modulation detection interference: across-frequency processing and auditory grouping,» *Hear. Res.* 79, 48-58.

**Yost, W. A., Sheft, S., and Opie, J.** (1989). «Modulation interference in detection and discrimination of amplitude modulation,» *J. Acoust. Soc. Am.* 86, 2138-2147.

**Young, E. D., and Sachs, M. B.** (1979). «Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers,» *J. Acoust. Soc. Am.* 66, 1381-1403.

**Zeng, F. G.** (2002). «Temporal pitch in electric hearing,» *Hear. Res.* 174, 101-106.

**Zeng, F. G., and Galvin, J. J., 3rd** (1999). «Amplitude mapping and phoneme recognition in cochlear implant listeners,» *Ear. Hear.* 20, 60-74.

**Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H.** (2004a). «On the dichotomy in auditory perception between temporal envelope and fine structure cues,» *J. Acoust. Soc. Am.* 116, 1351-1354.

**Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargava, A., Wei, C., and Cao, K.** (2005). «Speech recognition with amplitude and frequency modulations,» *Proc. Natl. Acad. Sci. U S A* 102, 2293-2298.

**Zeng, F. G., Popper, A. N., and Fay, R. R. (eds).** (2004b). Cochlear implants: auditory prostheses and electric hearing (Springer-Verlag, New York).

**Zwicker, E., and Schorn, K.** (1978). «Psychoacoustical tuning curves in audiology,» *Audiology* 17, 120-140.

**Zwicker, U. T.** (1984). «Auditory recognition of diotic and dichotic vowel pairs,» *Speech Communication* 3, 265-277.



Aurélie Zambeaux  
Communication et Marketing  
Tél: 01 41 88 01 59  
Email: ac@prodition.fr

1

**Oticon lance la première gamme d'aides auditives exclusivement destinée aux enfants malentendants : Oticon Safari**

Oticon lance Oticon Safari, première gamme complète d'aides auditives exclusivement conçues pour les enfants malentendants de la petite enfance à l'adolescence. Avec Safari, Oticon offre une gamme de fonctions hautement souhaitables qui ne sont actuellement pas disponibles dans les aides auditives du marché pédiatrique. Les nouvelles aides auditives conviviales pour les enfants possèdent une bande passante étendue pour supporter le développement du langage, un traitement binaural sans fil pour améliorer la conscience spatiale et les capacités de localisation, ainsi qu'une connectivité sans fil pour accéder facilement aux téléphones, télévisions et autres dispositifs de divertissement et de communication. Oticon Safari possède également un voyant LED intelligent unique intégré qui permet aux parents, enseignants et autres soignants de savoir avec certitude que l'aide auditive fonctionne correctement.



Crédit: OTICON



« Chez Oticon, notre philosophie People First s'est toujours centrée sur le développement d'aides auditives de haute qualité pour les plus jeunes et plus vulnérables membres de la population malentendante » a déclaré Christophe Aubert, Directeur Commercial Oticon. « Avec Safari, nous renforçons significativement notre engagement et nous

sommes les premiers à offrir une gamme spécialisée complète d'aides auditives exclusivement conçues pour les enfants. Nous savons que pour les audioprothésistes et leurs patients, une expérience réussie avec les premières aides auditives peut conduire à toute une vie de fidélité. Avec Safari, nous cherchons à gagner cette fidélité en offrant des aides auditives innovantes et conviviales pour les enfants pouvant évoluer avec eux, à travers leurs étapes développementales importantes en fournissant la fonctionnalité dont ils ont besoin pour atteindre leurs objectifs. »

### Accès total aux indices du discours pour une bonne acquisition du langage

Contrairement aux aides auditives traditionnelles qui limitent l'accès des enfants à la gamme complète d'indices de la parole, la bande passante plus large de 10 kHz de Safari offre des sons haute fréquence essentiels avec une audibilité cohérente et une meilleure précision jamais égalées.

### Accès aux sons de l'école... et de la vie moderne !

Safari, qui signifie « exploration et voyage » en Swahili, offre un grand nombre de fonctions de traitement sonore avancées que les audioprothésistes peuvent activer au fur et à mesure que les enfants franchissent des étapes développementales importantes.





Lorsque les enfants sont en âge d'aller à l'école, il est encore plus important d'avoir une interface facile avec les téléphones filaires et portables, télévisions, lecteurs MP3, ordinateurs, ordinateurs portables, jeux vidéos et autres sources de divertissement et sociales importantes. La connectivité sans fil de Safari offre un accès rapide et facile aux dispositifs de divertissement et de communication modernes en ajoutant l'accessoire élégant Streamer qui est équipé Bluetooth. Porté autour du cou, le Streamer ressemble à un lecteur MP3 et, il suffit d'appuyer sur un bouton pour qu'il transmette directement le son de sources audio externes aux aides auditives Safari.

### Des aides auditives qui ne sont rien qu'aux enfants

Solide et résistant, Oticon Safari est disponible en BTE et Power BTE traditionnels avec des piles 312 ou 13. Une vaste gamme de couleurs permet aux enfants d'exprimer leur personnalité et d'encourager un sens de l'identité et de la propriété. Des autocollants rigolos et des designs uniques, tels que les motifs zèbre et plante grimpante, permettent aux enfants de modifier leur look pour correspondre à leur humeur, leur tenue préférée ou une occasion spéciale.

Petit et ergonomique, Safari complète la gamme d'aides auditives à technologie avancée de pointe d'Oticon et les

programmes de support et de service pédiatriques. La nouvelle gamme Safari sera disponible à partir du mois de Décembre.

Pour davantage d'informations sur Oticon et sa nouvelle gamme Safari, consultez notre site Internet : [www.oticon.fr](http://www.oticon.fr) ou [www.oticonpaediatrics.com](http://www.oticonpaediatrics.com)

2

### Une nouvelle façon de penser avec : Modern Hearing Care



L'introduction d'aides auditives sans fil annonce un tournant significatif dans l'industrie de l'audition avec des conséquences aussi importantes que celles du passage du traitement analogique au traitement numérique. Pour les malentendants, ce changement représente la possibilité de participer plus pleinement à la vie moderne. Cependant, pour les professionnels pressés par le temps, de tels bonds en avant ajoutent une autre dimension à un marché déjà complexe. « Lors de moments charnières comme celui-ci, les audioprothésistes font face à deux défis » déclare Christophe Aubert, Directeur Commercial Oticon. « Ils doivent tout d'abord trouver les mots et les outils qui leur permettront de communiquer l'expérience de réelle connectivité à leurs patients et, de façon tout aussi importante, ils doivent imaginer comment appliquer cette tech-

nologie innovante lors de leur rendez-vous. »

Oticon est déterminé à aider les audioprothésistes dans leur quotidien et à leur offrir toutes les possibilités pour passer du concept à la réalité quotidienne. Ainsi, Oticon confirme son engagement de longue date à faire passer les besoins des patients en premier (People First). Oticon a donc décidé de développer le nouveau concept Modern Hearing Care (englobant l'ensemble des solutions ConnectLine disponibles pour les familles de produits Epoq, Dual et Vigo Connect). Un site est entièrement dédié à cette nouvelle façon de penser et il propose de nombreuses vidéos d'application, des interviews sur les perspectives du marché, des conseils d'audiologistes, des téléchargements pratiques sur les produits... « Avec cette initiative, il s'agit d'aider les audioprothésistes à faire face aux principales demandes de leurs patients lorsqu'ils recherchent une solution auditive et de se préparer à accueillir une nouvelle génération de patients exigeant des aides auditives s'adaptant aux technologies modernes ».

La gamme complète d'Oticon (Epoq, Dual et Vigo Connect) est désormais associée au système exclusif ConnectLine et offre aux audioprothésistes une opportunité inégalée de mieux faire face aux besoins de leurs patients avec des solutions intégrées simples pouvant optimiser l'utilité d'écoute, la facilité d'utilisation, le confort et l'esthétique.

Avec les solutions Modern Hearing Care, Oticon répond à cette demande d'être connecté au monde de communication moderne et satisfait les exigences actuelles des clients.

Pour davantage d'informations sur Modern Hearing Care et retrouver toutes les vidéos ou formations en ligne sur la gamme ConnectLine, visitez le site Internet : [www.modernhearingcare.fr](http://www.modernhearingcare.fr)





Simon Caillat-Thoranne  
Communication & Marketing  
04 26 23 22 03  
simon.caillatthoranne@unitron.com

## uHear étend son rayon d'action pour une meilleure sensibilisation à la perte d'audition

### L'auto-évaluation discrète et gratuite de l'audition désormais disponible en français sur iPhone™ et iPod touch®

**Kitchener, Ontario, Canada** - 7 janvier 2009 - Après le lancement réussi de uHear™, un test auto-administré de l'acuité auditive pour iPhone™ et iPod touch®, Unitron a annoncé aujourd'hui que l'application est désormais disponible dans le monde entier en français. uHear reste une des applications médicales gratuites les plus téléchargées sur iTunes. Jusqu'à présent, plus de 250 000 personnes ont profité de uHear et ont fait le premier pas vers une meilleure audition.

17 % (<http://www.hear-it.org/forside.dsp?forside=yes&area=34>) des adultes européens souffrent d'une perte auditive. Bien qu'une perte auditive ait un impact notable sur la qualité de vie, faire tester son audition peut rendre particulièrement anxieux et être, par conséquent, repoussé pendant de nombreuses années. Des études récentes montrent que certains adultes attendent plus de 10 ans (Données de l'étude Unitron réalisée du 8 au 14 avril 2009 par Angus Reid Strategies) entre deux tests auditifs. De plus, 13 % des répondants n'avaient jamais fait tester leur audition.

« La plupart des gens pensent que seuls les fans de musique Heavy metal, les travailleurs du bâtiment et les mineurs souffrent de pertes auditives, mais ce sont les choses du quotidien qui peuvent faire le plus de dégât. Le monde devient de plus en plus bruyant, mais augmenter le son n'est pas la solution » déclare Donald Hayes, Directeur de l'Audiologie Unitron. « En raison d'une peur, d'une confusion ou d'apathie, les gens ne font pas assez pour se prémunir ou mesurer une perte potentielle d'audition. Il n'y avait pas de moyen simple et individuel pour tester sa propre acuité auditive - jusqu'à aujourd'hui. »

Disponible sous la forme d'un test gratuit de dépistage de perte auditive, uHear permet aux utilisateurs d'identifier une potentielle perte auditive grâce à trois types d'évaluation : Sensibilité Auditive, Parole dans le bruit, et un Questionnaire sur les situations d'écoute courantes.

Facile à utiliser, uHear est conçu pour aider les utilisateurs à surmonter leur anxiété et sert de première étape vers l'identification et le traitement de la perte auditive. uHear offre également une fonction « Localiser » qui permet de trouver un professionnel de l'audition à proximité en cas de perte auditive détectée.

« Cette annonce illustre notre engagement profond d'accroître la sensibilisation à la perte d'audition. Une perte auditive a un effet considérable sur les interactions sociales et la perception de l'environnement. Elle reste cependant un aspect négligé de la santé globale. Avec cette nouvelle application en français, nous avons répondu à la demande du secteur en fournissant un outil d'évaluation discret sur plus de marchés, » déclare Cameron Hay, Président et Directeur Général d'Unitron monde. « Grâce au succès international de uHear et à ses téléchargements continus, nous permettons à encore plus de personnes dans le monde de tester facilement leur propre audition. » L'application en français pour iPhone™ et iPod touch® peut être téléchargée gratuitement sur l'iTunes store d'Apple : <http://itunes.com/apps/uhearfr>

### A propos de uHear™

Conçu par Unitron, uHear™ est un test d'audition qui vous permet d'évaluer votre audition au moment et à l'endroit que vous souhaitez. uHear vous aide à déterminer si votre audition se situe dans une plage normale grâce à un ou la totalité des trois tests suivants : Sensibilité Auditive, Parole dans le bruit, et un Questionnaire sur les situations d'écoute courantes.

### A propos d'Unitron

Unitron est un fabricant mondial innovant d'aides auditives haute technologie. Nous sommes profondément dévoués aux personnes concernées par une perte auditive et nous travaillons en étroite collaboration avec les audioprothésistes pour concevoir des solutions avancées, pratiques et accessibles à tous. Basée à Kitchener (Ontario, Canada), avec 16 bureaux dans le monde et des distributeurs présents dans plus de 53 pays, Unitron, membre du groupe Sonova, travaille en étroite collaboration avec les audioprothésistes pour répondre aux besoins de leurs clients. Pour plus d'informations, visitez le site [unitron.com/fr](http://unitron.com/fr).

1

## Compte-rendu de l'EPU Année 2009

Nouveau record battu avec plus de 850 personnes présentes à l'EPU 2009 organisé comme chaque année par le Collège National d'Audioprothèse à la Cité des Sciences à Paris. Le sujet qui faisait suite aux conférences de l'année 2008 sur « Psychoacoustique et Surdit   » avait d  j attir   beaucoup de monde l'ann  e derni  re (plus de 750 personnes). L'objectif fix   cette ann  e   tait pour chaque orateur de faire un pont entre la pr  sentation qu'il avait faite l'ann  e pass  e et le sous-titre de cet EPU 2009 : « Application Proth  tique ». Il est toujours difficile de s'attaquer au c  ur de la culture technique et scientifique d'un m  tier car elle est faite de certitudes, d'habitudes, de croyances et de connaissances. Ce m  lange n'est pas toujours simple    remettre en cause. Mais, l'objet de journ  es comme celles-ci reste de remettre un peu d'ordre en essayant de repousser les certitudes et les habitudes qui ne sont pas toujours tr  s utiles, de marginaliser les croyances et d'accro  tre les connaissances. Ce n'est pas toujours facile parce que la psychologie   tant alors de la partie toutes les r  sistances   mergent ; la consigne devient alors pour tout intervenant de le savoir et d'expliquer avec tact et d  termination son point de vue. Ce n'est pas toujours simple. Nous allons pr  senter ici non pas un compte rendu in-extenso de ces journ  es dont les conf  rences seront publi  es mais plut  t certaines questions de fond qui n'ont pu   tre pos  es faute de temps.



Apr  s une introduction du Pr  sident **Eric Bizaguet**, qui a rappel      l'auditoire le r  le du Coll  ge National d'Audioproth  se et son implication dans l'enseignement, **A. Coez** nous a r  sum   en 45 minutes 13 ou 14 heures de conf  rences faites en 2008. Ce rappel a conduit l'auteur    souligner quelques points importants qui doivent   tre connus et compris de l'auditoire pour mieux int  grer les techniques proth  tiques et leurs bases (voir les Cahiers de janvier 2009 pour le r  sum   de l'EPU 2008).



Deux conf  rences : **Frank Lefevre** « mesures vocales dans le silence et dans le bruit » L  on Dodel   « localisation sonore spatiale,   quilibrage sonore et st  r  audiom  trie ». Ces deux conf  rences n'ont pas eu lieu le m  me jour mais elles fonctionnaient assez bien ensemble.



La premi  re r  alis  e par Frank faisait   tat d'exemples sonores, en fait si    mes yeux et surtout    mes oreilles, elle a pris une grande importance c'est que son auteur a essay   d'amener l'auditoire    structurer son savoir sur un sujet qui est aujourd'hui mieux d  fini qu'on ne le croit. Il a pr  sent   un travail de synth  se essentiel sur les mesures r  alis  es    partir de signaux de parole que ce soit en audiom  trie vocale lexi-

cale ou non-lexicale dans le silence et dans le bruit. Le manque de temps n'a pas cependant pas permis d'aborder une question essentielle pour nous qui touche tout autant le travail de **L  on Dodel  ** et qui est li      la notion d' « entropie » que nous expliciterons tr  s bri  vement ci-dessous. Frank a montr   que bien conduite l'audiom  trie vocale pouvait s'apparenter en quelque sorte    une mesure de coh  rence en acoustique (qui repr  sente rappelons-le la mesure de la somme des distorsions). L  on a montr   que l'  quilibrage par les tests vocaux en milieu bruyant ne devait pas   tre con  u comme une option mais bien comme l'appareillage st  r  ophonique : une r  gle. La conception de ces tests comme nous allons essayer de le pr  senter sugg  re qu'ils ont en th  orie et en pratique une puissance qui leur conf  re un int  r  t tout particulier dans le cadre de la r  habilitation proth  tique. Les principaux tests ont   t   pass  s en revue. Il faut n  anmoins rappeler pr  alablement au lecteur qu'il doit savoir utiliser l'audiom  trie vocale comme cela se fait dans les standards internationaux.

**1/ Audiom  trie Vocale Lexicale (AVL) d'identification (AVLi) : elle vise    mettre en   vidence le niveau de l'intelligibilit   par rapport au Seuil de R  ception Vocale (SRV).**

**Objectif : informer sur la qualit   de reconnaissance de la parole en 2 ou 3   tapes.**

**a/ La d  termination du Seuil de D  tection de la Parole (SDP ou, chez les anglo-saxons, SDT Speech Detection Threshold) :**

Peu utilis   en pratique, il peut se faire    la voix au casque en utilisant une r  p  tition de syllabes bi,bi,bi... A partir d'un niveau tel que le sujet ne reconna  t pas qu'il s'agit de sons de parole jusqu'au moment o   il signale qu'il ne sait pas ce que vous dites mais qu'il lui semble qu'il s'agit de mots... le SDT est   -peu pr  s    5 dB HL chez le sujet normal.

Il est couramment admis (dans le monde anglo-saxon de l'audiologie) d'utiliser le SDT dans la limite de personnes ne comprenant pas très bien ce que l'on attend d'elles, par exemple des enfants polyhandicapés, et ce pour essayer de déterminer un seuil de sensation. Ce test est peu utilisé.

### b/ Recherche du seuil de réception vocale (SRV ou SRT anglo-saxon) :

Ce test correspond au niveau précis (à 2dB près) auquel on a 50% de reconnaissance de mots dissyllabiques. Ce test se fait au casque en audition monaurale. Dans le cadre de L'Audiométrie Vocale Lexicale (AVL) on utilise le plus souvent le test de Fournier ou le test de Lafon (mots dissyllabiques). L'objectif : rechercher une cohérence entre audiogramme et audiométrie vocale à partir de critères simples (**cf conférence de J. Jilliot et col. EPU 2007**).

La valeur retenue correspond à **50% de reconnaissance de mots dissyllabiques** dont les 2 syllabes ont à peu près le même niveau acoustique moyen et la même dynamique. Ce seuil présente la caractéristique de correspondre à 10dB près à **la moyenne arithmétique de la somme des pertes tonales 500 + 1000 + 2000Hz/ 3 = N - 2 dB** (selon Carhart & Porter 1971) **si il n'y a pas d'écart trop important entre les fréquences entrant dans le calcul en terme de niveaux**. Ainsi si les 2 premières fréquences sont à un niveau de 20 dB et la troisième à 70 dB on ne vérifiera pas la règle de la somme des 3 valeurs mais simplement des 2 premières (500 et 1000Hz). (Noter un écart normal de 10 à 12 dB entre ces 2 niveaux, ou, chez nous : SDT-SRT).

*Remarque à propos de la passation rapide du test SRV : elle peut être accélérée. Ainsi avec le Fournier dès lors que 5 mots seront reconnus sur une liste de 10 on pourra descendre le niveau de 5dB puis on pourra réduire le pas à 2 dB etc... (Réciproquement pour monter). Le SRV est à-peu-près à 15 dB chez le sujet normal ce qui deviendra la valeur de référence pour la suite.*

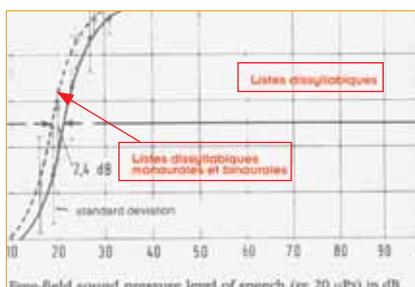
On peut aussi avoir une interprétation plus psycho-acoustique de l'évolution des courbes de reconnaissance à des niveaux supra-limaires.

## 2/ Audiométrie vocale lexicale de sensation, structuration et stratification du champ auditif : identification et fonctions P/I de référence.

Il y a une autre application des fonctions possibles des tests vocaux, c'est le traçage de fonction performance/intensité. Les profils sont bien connus pour les mots dissyllabiques ainsi que pour les mots monosyllabiques phonétiquement équilibrés (voir les **fonctions de référence ci-dessous**) :

### a/ Les listes dissyllabiques.

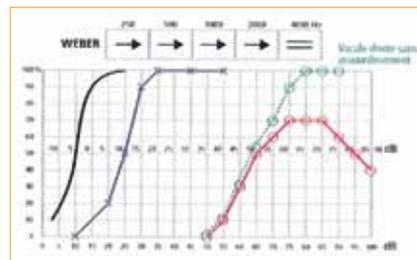
L'objectif du test : comptabiliser les mots erronés en soumettant le sujet à une série d'éléments de listes de mots dissyllabiques en augmentant le niveau d'émission de manière régulière jusqu'à obtenir une reconnaissance maximale (qui ne sera pas nécessairement de 100%. Au-dessus du seuil à 50% de reconnaissance (SRV) la progression de reconnaissance des mots est de 50% pour 5 dB de variation de dynamique c'est-à-dire 10% de progression par dB. Ce test n'est donc pas forcément idéal du fait de sa faible sensibilité (montée trop rapide) pour tester l'intelligibilité en dehors de la notion de seuil. Rappelons que cette mesure prend théoriquement tout son sens dans le cadre de l'évaluation de la cohérence dans la confrontation des résultats vis-à-vis de l'audiométrie tonale liminaire. **Il est donc peu intéressant de le pousser au-delà du SRV (c'est-à-dire 50% de reconnaissance des mots) si l'on souhaite réaliser un test d'intelligibilité. Par contre le raisonnement peut être différent si on souhaite tracer une fonction P/I sous la réserve de ne pas arrêter là l'audiométrie vocale puisqu'il ne s'agit pas alors d'une mesure fine de l'intelligibilité (10% de progression/dB).**



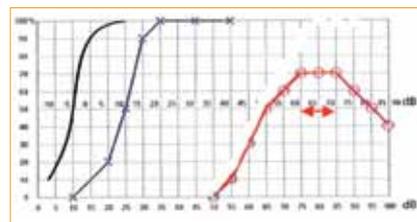
Une recommandation : prendre garde de ne pas aboutir à des résultats du

type de celui présenté par le Dr Martine Ohresser dans l'excellent recueil des audiométries de la Société Française d'Audiologie, en étant attentif au fait qu'il conviendra de masquer l'oreille non testée si :

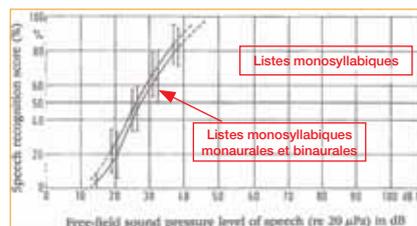
SRT - TTC (Transfert Trans-Cranien) ≥ CO Moyenne controlatérale



Il faut noter de plus qu'une courbe comme la courbe rouge présentée ci-dessous **avec un test dissyllabique** suggère la présence d'une limite dans la capacité de reconnaissance de la parole ainsi que la présence d'un recrutement. Dans le cadre d'un test phonétiquement équilibré (Type Lafon), ce qui n'est pas le cas ici si on se réfère à la fonction de référence, un tel résultat doit conduire à une évaluation particulière permettant d'opposer la stabilité de reconnaissance vocale maximale (ici 10 dB) à un indice de stabilité Is. Cette démarche permettra de s'interroger sur la présence d'un recrutement (le score n'atteint pas 100% de reconnaissance) ou sur un problème dit « d'intégration ».



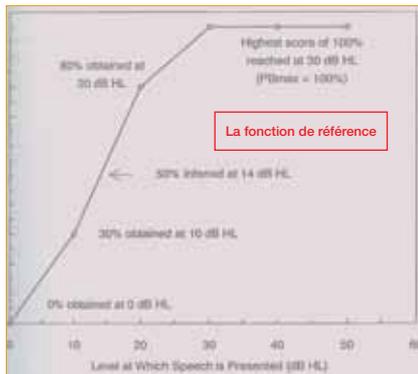
### b/ Les listes monosyllabiques :



L'objectif est de comptabiliser non pas les mots inexacts mais les phonèmes erronés, (équivalent chez nous : le test de Lafon). **Ces tests sont monosyllabiques** et phonétiquement équilibrés. Ce test démarre à-peu-près au même

niveau que le test dissyllabique c'est-à-dire vers 10 dB pour atteindre 50% de reconnaissance vers 25 dB et 100% vers 45 dB. La croissance du taux d'identification des phonèmes est plus lente que pour le test **SRV (10% par dB)** il est de **2,5% par dB** ce qui en fait un test plus sensible. **Le profil de la fonction P/I sera donc plus incliné que pour le test dissyllabique.**

**Les repères :** le profil de croissance de la fonction normale doit être connu de l'audiométriste (cf Fig. ci-dessous)



La méthode d'obtention est simple, il faut entre 7 et 10 éléments de liste émis à des niveaux croissants. **Le profil de la courbe est une donnée importante** pour l'audiométriste.

**Les points spécifiques :** si cette valeur n'atteint pas 100% de reconnaissance il convient alors d'augmenter le niveau par pas de 5 ou 10 dB, si la réponse plafonne en dessous de 100% on notera la valeur. Puis, par une approche assez simple mais très efficace on recherchera **le niveau de confort vocal puis le maximum vocal acceptable** (Critères de De Bock-Renard) en augmentant le niveau des mots jusqu'à ce que le sujet soit gêné. A ce niveau on fera une mesure de reconnaissance vocale et on vérifiera si la performance a baissé. Dans ce cas on revient 5 ou 10 dB en arrière. Il convient toujours **de rechercher si nous pouvons disposer de plus ou de moins de 30 dB de stabilité en termes de reconnaissance vocale optimale.**

**L'indice de stabilité (Is) dans la pratique :** si rapidement la performance décroît il faudra poursuivre par bonds de 10 dB jusqu'à un niveau (si possible supérieur ou égal à 30 dB). On peut alors évaluer un « **indice de stabilité** » que l'on calcule comme suit :

$$Is = \frac{\text{Score Max} - \text{Score Min}}{\text{Score Max}}$$

Si le résultat est inférieur à 0,25 il n'y a pas de problème rétro-cochléaire. Par contre si le résultat est entre 0,25 et 0,45 on peut en suspecter un (résultat à confronter aux tests dits de recrutement). Notons par ailleurs qu'une baisse de la performance avec les mots dissyllabiques suggère la présence d'un recrutement. Dans le cas d'un test monosyllabique, la reconnaissance n'atteint pas 100% mais reste stable.

### 3/ L'audiométrie vocale non lexicale (AVNL) et la structuration du champ auditif : intérêt, identification et fonction P/I.

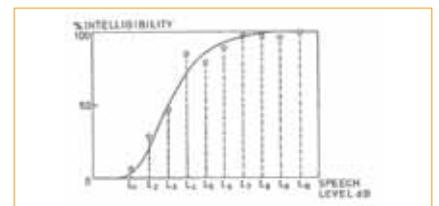
#### a/ Discussion préalable :

Nous avons présenté de manière très succincte l'approche la plus classique de l'audiométrie vocale. Il semble cependant qu'aujourd'hui l'utilisation de tests dits d'intelligibilité soit non pas remise en question bien au contraire, mais discutée pour en cerner les limites en particulier lorsqu'il s'agit de tester strictement des capacités de décodage chez un sujet normo-entendant ou malentendant. Parmi les questions qui se posent il y a bien sûr les biais tels que la connaissance du lexique, la fréquence des mots etc... qui sont susceptibles d'entraîner des performances qui, en plus de l'imprécision liée au nombre d'items par élément de liste, introduisent un autre biais qui est celui lié à la connaissance de la langue car elle conduit le sujet à faire des assemblages de phonèmes dont la probabilité dépend de la connaissance qu'il a de sa propre langue plus que de ses capacités réelles de décodage. Les résultats sont alors rapidement dépendants de facteurs autres qu'acoustiques. Une analyse plus rigoureuse de la construction des tests montre qu'en réalité quelle que soit la forme retenue même dans les tests de mots sans signification on retrouve toujours des assemblages plus probables que d'autres vis-à-vis d'une langue donnée.

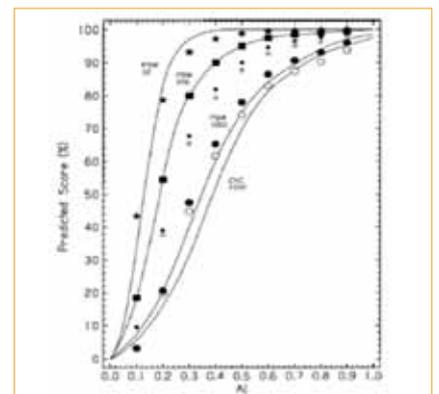
Pour sortir de ce paradoxe il faut alors poser le problème différemment. On peut ainsi poser la question **non plus en termes d'identité de fréquence entre la répartition des phonèmes d'un test et ceux de la langue du test, mais en termes d'égale occurrence d'assemblage d'un phonème avec un autre dans un**

**arrangement.** Cette approche présente un intérêt acoustique qui ne déborde pas sur des problèmes de mémoire ou d'articulation (assemblage de phonèmes difficile à prononcer ou à entendre (ex : « trg » mot sans signification pour nous ; la rue en Yougoslave). Dans ce contexte maîtrisé et en choisissant opportunément certains arrangements, il devient possible de concevoir des tests composés de syllabes CV, VC, de mots CVC, de mots VCV... qui vont présenter le maximum de désorganisation (d'entropie) c'est-à-dire mélangeant le phonétique, le phonémique (*un phonème en début et en fin de mot n'ayant pas exactement les mêmes caractéristiques*), le phonologique etc... Dès lors, il devient inutile de se poser des questions vis-à-vis du sens ou de la suppléance mentale. La règle d'équiprobabilité élimine la question et va conduire à intégrer tous les possibles pour accroître le calcul du probable. Une fois ces tests construits on peut concevoir par exemple 2 types de passation : sous forme de fonction performance-intensité ou bien sous forme de test de reconnaissance avec un niveau de bruit variable pour tester la performance de reconnaissance vocale en fonction du rapport S/B P/ (S/B) (cf ci-dessous les résultats pour le sujet normal).

#### b/ Le principe de report des données (cf Fig. ci-dessous) :



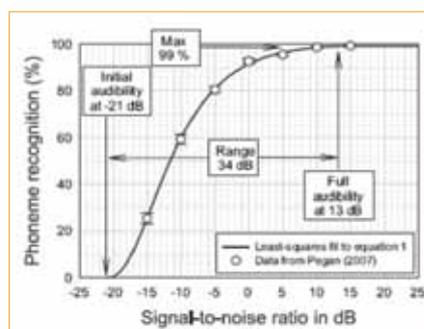
Les repères pour la fonction de référence par rapport aux tests AVL :



A l'observation de la figure ci-dessus on peut noter que la difficulté supposée d'un test de mots sans signification n'est pas justifiée et que les résultats sont statistiquement tout à fait comparables à ceux d'un test de mots du lexique phonétiquement équilibré.

**c/ 2 représentations intéressantes :**

- la représentation sous forme d'une fonction de sensation classique P/I partant de 0% de reconnaissance de la parole pour aller vers le maximum de reconnaissance vocal,
- une représentation prenant en compte la performance en fonction du rapport signal/bruit. (Cf ci-dessous fonction P/s/b chez un sujet normal).



De manière générale on retiendra qu'avec un rapport signal / bruit de -20 ou -25 dB la performance dans la reconnaissance chez un tel sujet est de l'ordre de la chance. Cette approche est très intéressante car elle permet d'évaluer l'émergence d'une réponse, ce niveau correspondant au seuil en dessous duquel les bonnes réponses correspondent à la chance c'est-à-dire qu'elles sont en dehors de toute valeur statistique. Cet aspect semble à lui seul déjà présenter un intérêt important du point de vue de l'analyse et nous le retiendrons comme un élément fort de l'approche par les tests de mots sans signification (Tests de F. Lefevre ATEC ou de Léon Dodelé).

*Remarques : L'utilisation du terme : « mots sans signification » devrait être modifiée car dans toutes les listes il y a des syllabes qui appartiennent à la langue qui peuvent être signifiantes ( ex. ta, ma, ni, nu...). La rigueur voudrait que le test soit composé de mots dont chaque « phonème » ou « phone » a une égale probabilité d'occurrence. Dans ce cas on parle alors d'ensemble de syllabes ayant une entropie maximum.*

**\*Remarque complémentaires à propos de l'entropie**

Si P<sub>n</sub> est la probabilité d'un évènement n parmi N possibles alors la densité d'information :  $\ln = \log^2 (1/P_n)$  avec n = 1, 2,..., N (rappels :  $\log_b a = \text{Log } a / \text{Log } b$  de plus, lorsque b = e ;  $\log_b a$  s'écrit Log a)

Entropie  $H = \sum_{n=1}^N P_n \cdot \ln$

$PN = \sum P_n$

**d/ La mise en pratique :**

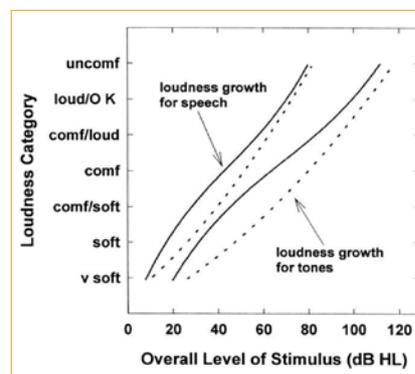
Ces tests de mots sans signification permettent de passer d'une évaluation clinique - qui finalement ne se passait pas si mal avec les tests classiques que nous avons évoqués plus haut - à des outils auxquels on peut - comme l'a montré Frank Lefèvre - **appliquer une puissance de calcul qui les rend incomparables sur le plan opérationnel dès lors que l'entropie est maximum.** Il est par ailleurs tout à fait possible d'en accroître encore la puissance en les utilisant dans le bruit et dans ce domaine l'« onde vocale globale » de Léon semble être un instrument intéressant. Pour ne reprendre qu'une remarque faite par Frank, le fait de pouvoir traiter les réponses en introduisant le principe de « **seuil de chance** » permet de dissocier les réponses statistiquement significatives de ce qui relève justement de la chance. Cet aspect représente un progrès qui évite de prendre les apparences de certaines réponses pour la réalité au sens statistique du terme (voir figure ci-dessous).

*Sur la figure ci-dessous, on remarque qu'au fur et à mesure de l'amélioration du rapport S/B la reconnaissance s'améliore. Elle devient assez bonne pour un rapport S/B entre -6 et 0 dB ; bonne pour un rapport de 0 à 6 ; excellente au-dessus.*

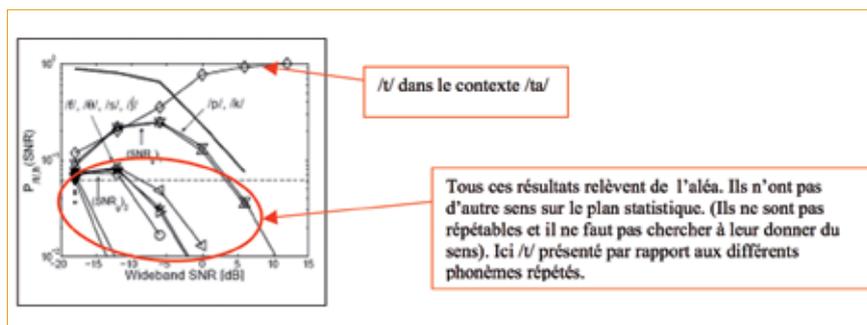
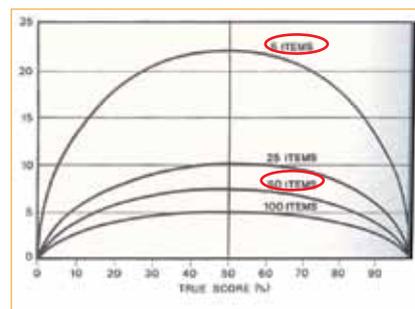
Une analyse de ce type permet entre autres de développer des études qua-

litatives sur des thématiques telles que l'effet de la compression et l'émergence de nouvelles confusions phonétiques (Voir figure ci-dessous pour exemple).

Un autre aspect de la réalité vocale est présenté dans la figure ci-dessous :



On remarque sur cette figure que les fonctions de sensations vocales et tonales ne se recouvrent pas totalement. C'est un aspect important qu'il faut aussi rapprocher d'un point important soulevé par Frank, c'est-à-dire la notion d'incertitude liée au nombre d'items.





Deux autres conférences complémentaires l'une de l'autre : **Eric Bizaguet** et **Yves Lasry**. Nous reprendrons les sous-titres : « Protocole initial et organisation des tests complémentaires » pour le premier (dont la conférence a été faite après par souci de clarté) et, « mesures psycho-acoustiques tonales » pour le second. Il s'agit là du deuxième sujet majeur de la matinée puisque c'est tout simplement notre quotidien qui est en question au travers d'un outil remarquable qu'est en train de concevoir Yves Lasry à la demande du Président du Collège et de Xavier Renard qui travaille beaucoup avec Yves à la définition des tests et à leur usage rigoureux. Je suis face à mon patient, je viens d'effectuer une audiométrie classique après avoir lu le compte rendu du médecin et, première surprise, Eric nous suggère de n'interroger le patient qu'une fois cette audiométrie réalisée car déjà nous avons pu confirmer et peut-être deviner un certain nombre de difficultés que nous allons rencontrer au travers du comportement du patient lorsqu'il a répondu aux tests. Cette approche est à mon sens intéressante dans la mesure où elle est en parfaite harmonie avec l'outil informatique mis au point par Yves. En effet elle va induire un contenu relationnel entre le patient et le prothésiste qui va pouvoir confirmer l'audiométrie médicale et permettre de passer ainsi à la partie prothétique proprement dite avec une liste de tests complémentaires accessibles par l'outil informatique dont la liste des tests, leur précision et leur souplesse d'utilisation ont de quoi laisser pantois le plus exigeant des professionnels de l'audiologie. Rappelons que les tests présentés comporteront un Bekesy, tous les tests de recrutement, des tests de mémoire immédiate etc... S'y ajouteront les courbes d'accord, les tests de masquage et tous les tests susceptibles de trouver une application en audiologie prothétique. Nous n'in-

sisterons pas plus sur ces deux conférences pour ne pas dévoiler le contenu exact de l'outil qui sera présenté dans la version définitive au moins pour trois modules de tests, au congrès de mars 2010 à Paris et donnons rendez-vous aux professionnels pour non seulement le découvrir mais aussi le tester et en faire l'acquisition.



Les conférences de **Stéphane Laurent** et **François Degove** ainsi que celles de **Christian Renard** et **Jean Baptiste Delande** (nouveau Collégien). Quatre conférences sur des sujets délicats : les compressions, les canaux, les taux, les seuils d'enclenchement etc... suivie de la conférence de Christian sur les débruiteurs, les détecteurs de pics et autres rehausseurs enfin, la conférence de J. B. Delande nouveau venu au Collège sur les microphones directionnels. Ces conférences étaient en fait indissociables. Elles représentaient un tout car le rehaussement et la compression sont bien dans la même idée que le fonctionnement de la membrane basilaire qui tout à la fois comprime, et amplifie de manière sélective jouant ainsi le rôle de détecteur de pics. Nous avons donc partagé une petite partie les uns avec les autres mais avec une conception qui devait permettre aux uns et aux autres d'avoir une vision d'ensemble de sujets pour le moins difficiles. Notez que si ils étaient simples cela se saurait. Ainsi, l'interférence entre les jeux des microphones et la compression, de même que les systèmes de débruitage, sont d'une grande subtilité même si les nouveaux systèmes d'analyse des prothèses auditives permettent de travailler de manière élective sur chacun d'entre eux, il va de soi que lorsque les aides auditives travaillent dans les oreilles des patients, les uns et les autres se mélangent. C'était un sujet très vaste et sans doute trop vaste pour le temps imparti. Mais telle est la règle et il faut l'accepter.

Le parti pris : Stéphane Laurent faisait une présentation sonore « imagée » François abordait les sujets de manière plus théorique. faisait une présentation sonore « imagée » François abordait les sujets de manière plus théorique. Cette double approche renvoyait à une vision que nous avons l'un et l'autre du sujet que nous enseignons tous les deux : c'est un sujet difficile où règne essentiellement confusion et a priori.



Donc 1, il faut mettre du concret sous les mots 2, il faut mettre de la rationalité et de la connaissance derrière les concepts. Il est difficile de présenter les images sonores de Stéphane mais elles exprimaient très bien ce qu'est une dynamique, une compression de dynamique, un seuil et un temps de compression etc... Les esprits devaient donc être prêts pour la suite plus abstraite. Pour François Degove, la première mission était de rappeler l'organisation anatomo-physiologique qui sous-tend les mécanismes de non-linéarité qui caractérisent une oreille normale puis, de rappeler où se situent les pertes de

non-linéarités. Nous avons donc rappelé le rôle des différents groupes cellulaires impliqués dans ces mécanismes et en particulier le rôle central des CCE ainsi que des CCI tout en essayant de rap-peler les données que nous pouvions tirer de ces informations. A partir de là nous avons présenté des systèmes de compression qui reprennent certaines caractéristiques pour étayer une démar-che qui allie connaissance de la physio-logie et cohérence dans les choix tech-nologiques. Nous avons par exemple essayé de montrer comment concevoir des circuits mimétiques reprenant les caractéristiques temporelles liées aux mécanismes particuliers du Complexe Olivaire Médian pour favoriser le pas-sage des transitoires dans le bruit. Dans un deuxième temps nous avons montré pourquoi et comment une normalisation de sonie peut venir contrarier la perte totale ou partielle des mécanismes de suppression indispensables à l'émer-gence de l'information au niveau de la cochlée et donc du codage neural. Nous avons ensuite présenté un type particu-lier de compression qui permet d'éviter la contradiction qu'il y a entre normali-sation de sonie et respect du contraste -élément essentiel de la perception- comme nous l'a rappelé Paul Avan lors de l'EPU 2008. Notons au passage que s'il apparaît nécessaire de ne pas avoir trop de canaux de compression (au-de-là de 8 il semblerait que cela dégrade certains aspects du signal en particulier du signal vocalique) comme le suggè-rent la plupart des études aujourd'hui, il faut comprendre cette restriction avec un peu de recul, ce qui n'a d'ailleurs pas échappé aux fabricants qui les regroupent subtilement la plupart du temps malgré un nombre important annoncé. En fait il ne faut pas faire la confusion entre canaux de compression et canaux d'amplification. Le troisième volet pré-sentait des résultats de recherches qui prennent un essor important depuis quelque temps pour tester et dévelop-per des processeurs dont la stratégie vise à corriger les défauts enregistrés au niveau des réponses neurales en simulant des pertes liées à des déficits de CCE ou bien de CCI (resp. perte neu-rale) ou bien des 2 groupes cellulaires, par exemple dans les atteintes de la strie vasculaire en question que nous avons évoqué plus haut. Ce travail a

été l'occasion de montrer qu'il y avait un certain nombre de techniques de modélisation qui aideront à poser des questions conceptuelles nouvelles dans les années à venir (seuil de saturation comparé au niveau d'inconfort, com-pressions à seuils et taux multiples qui sembleraient être une meilleure réponse à la fonction de sensation telle qu'on l'a présenté en 2008 -Inex- qui est une fonction incurvée négativement entre 40 et 70 dB et qui serait apparemment plus représentative de l'évolution de la sensation dans cet espace de la dyna-mique auditive). Il y a là par exemple un point à mieux comprendre. En réalité cette approche pour théorique qu'elle apparaît ne reflète qu'un maniement adroit et raisonné de la compression et un choix judicieux du type de com-pression. Il ne fait aucun doute que ce chapitre prendra de l'ampleur dans les années à venir. Dans la partie 2 nous avons abordé les techniques permettant de définir les zones et les canaux d'am-plification par opposition aux zones de non amplification essentiellement liées aux pertes totales de CCI et aux pertes neurales associées. Nous avons soulig-né l'importance qu'il y a pour nous à progresser dans leur définition et l'ac-tualité de cette question au travers de la réparation cellulaire qui pointe son nez. Ensuite, nous avons rappelé l'intérêt qu'il y a à bien explorer la fonction pour en tirer des informations stratégiques à partir d'un exemple de « softness imperception », concept qui a été décrit lors de l'EPU 2008 et au travers duquel nous avons essayé de montrer que le choix des paramètres de correction ne pouvait résulter d'une application systé-matique d'une compression type WDRC à des individus qui présentent une gêne importante liée au niveau sonore. Ce point nous amène d'ailleurs à rappeler que dans bien des cas l'utilisation de questionnaires élaborés (APHAB etc...) permettrait de mieux faire exprimer aux gens ce qu'ils veulent dire plutôt que de s'appuyer sur ce que nous croyons entendre. La troisième et dernière partie traitait de l'application et donc du calcul proprement dit des taux et des seuils de compression au travers de 2 logiciels que nous avons mis au point avec des étudiants.

En fait, la difficulté que nous avons ren-

contrée dans la présentation de la com-pression est que celle-ci, lorsqu'elle est vue à partir d'une aide auditive, apparaît comme un simple instrument au servi-ce du confort. Si la finalité du travail de compression cochléaire semble cohé-rente avec ce point de vue, l'activité de régulation des signaux présents simul-tanément dans la cochlée semble s'ap-puyer sur un mécanisme plus subtil. Le rôle de la compression cochléaire serait d'ailleurs presque contre-productif si, parallèlement, l'une des fonctions essen-tielles de l'organe de Corti (membrane Basilaire + CCE) n'était pas celui dévolu au masquage suppressif dont le but est en fait le nettoyage du paysage autour des pics de résonance les plus francs pour permettre au système auditif de les saisir et de s'auto-organiser autour sans être perturbé par certains sons simul-tanés présent autour du signal le plus proéminent. Ce travail semble nécessi-ter au passage une grande finesse de la régulation du niveau d'amplification par l'intermédiaire de la suppression pour compenser d'éventuels effets négatifs de la compression !

Pour revenir à la conférence de Chris-tian il avait à faire passer, entre autres, le message de l'impossibilité de réali-ser des mesures de gains prothétiques avec des aides auditives munies de ces systèmes détecteurs de bruits du fait que ceux-ci baissent systématiquement leur gain sur des signaux de type sons wobulés ou sons purs. Noter qu'il en va de même pour les mesures in-vivo car les aides auditives récentes détec-tent très rapidement les signaux qui ne sont pas de la parole y compris les signaux ILTASS et, de la même manière, atténuent les réponses. Ces particulari-tés sont importantes surtout lorsqu'on travaille avec des enfants car lorsqu'ils sont petits et ne possèdent pas le lan-gage, il est très difficile de recouper des réponses entre elles. Ajoutons à propos de l'enfant tout petit que le conseil est aussi tout simplement de ne pas utiliser les débruiteurs pour les habituer à traiter paroles et bruits et à extraire l'informa-tion de cet environnement. On en revient à la dernière conférence car les jeux de microphones seront les seuls systèmes que l'on s'autorisera à utiliser dans l'ap-pareillage de l'enfant pour favoriser le rapport signal/bruit. Les dernier systè-

mes décrits sous la forme de rampes de microphones présentent de ce point de vue des effets très sensibles mais il faut rappeler que leur entretien rigoureux est impératif faute de quoi les résultats seront très fortement dégradés.

### Le samedi

Le premier sujet traité : transposition et compression de fréquences. Sujet d'actualité traité par **Arnaud Coez**. Ce sujet est évidemment à la croisée de chemins de l'audioprothèse puisque il nécessite d'une part un traitement très rapide du signal pour capter, compresser et restituer le signal de sortie après un temps de l'ordre de quelques millisecondes et, d'autre part, une approche clinique impliquant entre autres une très bonne maîtrise de l'audiométrie et du test de Moore.



Dans sa présentation Arnaud a d'abord bien expliqué la différence entre transposition fréquentielle et compression fréquentielle. C'était un préalable important car les deux systèmes ambition-

nent un supplément d'information jugé essentiel pour la perception en particulier dans le bruit, mais en prenant des directions tout à fait différentes sur le plan technologique. Il faut rappeler que ces systèmes ne sont pas récents. Il suffit de rappeler la prothèse de Lafon qui plutôt qu'une prothèse à transposition était une forme de vocoder qui permettait de capter une ou deux bandes de fréquences et de générer des réponses modulées dans des bandes plus basses. Un autre système étant le playback à vitesse ralentie. un troisième, transposer une bande HF dans une zone plus grave « par-dessus » le signal situé en dessous de la fréquence de coupure. La compression de fréquences non-linéaire (les fréquences les plus hautes étant compressées plus que les HF moins hautes. Abaissement de fréquences linéaire. La revue de la littérature fait apparaître une amélioration intéressante pour la transposition dans certains cas de zone morte et pour les fricatives. Une autre étude faisant apparaître une amélioration chez l'enfant dans les cas de surdités sévères profondes. Pour ce qui concerne la compression fréquentielle, une amélioration serait présente pour les surdités moyennes à profondes. Si cette approche paraît intéressante elle semble aussi trouver des limites avec les surdités à très forte pente en particulier dès lors que la fréquence de coupure s'approche de 1KHz. Il y a à n'en pas douter un gros effort à faire pour améliorer ces dispositifs qui sont de toute évidence un progrès dans l'appareillage. Notez au passage que leur adaptation devrait s'accompagner de contrôles systématiques avec des tests sans signification dans le silence et dans le bruit. Il faudra à ce sujet un jour ou l'autre que le Collège préconise, statistiques à l'appui, des protocoles pour ce type d'appareillage. Notez aussi qu'il semble que l'entraînement auditif apporte une amélioration sensible des résultats.

Le deuxième sujet : Autres traitements spécifiques du signal. Anti-larsen et communication entre les deux appareils» **Jérôme Jilliot**. Jérôme n'est plus un jeune Collégien même s'il est jeune et cela se sent bien lorsqu'on écoute ses conférences.

Il connaît parfaitement les outils qu'il utilise que ce soit les aides auditives ou les instruments de mesure. Il a bien montré entre autres qu'il faut se méfier des anti-larsen en particulier lorsqu'on les utilise avant d'avoir effectué certains réglages sur les HF. Pour ce qui concerne la communication entre les appareils certains travaux récents tendent à en montrer l'intérêt, et, en particulier, l'amélioration dans l'équilibre entre OD/OG favorise une meilleure perception dans le bruit. Là aussi l'utilisation de tests de mots sans signification avec et sans bruit devrait permettre une amélioration objective de leur intérêt. Ce serait une raison suffisante pour que le Collège édite non seulement les tests mais aussi les outils d'analyse qui vont avec.

La troisième conférence : « Adaptation prothétique et contrôle d'efficacité. Interférences entre réglages, etc... **S. Garnier** et **F. Lefevre**. La thématique de ce sujet était en fait de reprendre un certain nombre de résultats des conférences précédentes. Nous l'avons déjà souligné dans le compte rendu des conférences du vendredi : le découpage est artificiel. Il faut bien se faire à l'idée que rehaussement et compression sont les deux versants d'une même problématique : le contraste que le sujet normal restitue comme nous avons essayé de le montrer à partir de calculs simples en utilisant le masquage suppressif qui malheureusement disparaît avec les surdités qui impliquent des pertes substantielles de CCE. Dans ce cadre le travail de paramétrage n'a évidemment de sens que lorsque le prothésiste a une vision assez claire des groupes cellulaires impliqués dans la perte auditive pour ne pas se tromper de cible. C'est tout un art qu'il n'est évidemment pas simple de mettre en lumière en particulier si on n'a pas d'informations suffisantes à propos de l'état des lieux cochléaires.

La quatrième conférence : multiprogrammation, mesures psycho-acoustiques et traitement du signal. Auteur **Bernard Hugon**. Le résultat d'un appareillage dépend bien sûr du travail de réglage du programme de base de la prothèse, mais aujourd'hui ce n'est plus le seul et les fabricants ont bien compris l'intérêt du multiprogramme.

Mais comment faut-il régler les programmes 2, 3 etc ... A cette question l'auteur a d'emblée souligné que les programmes dits supplémentaires étaient en fait des programmes à part entière et non des sous-programmes. Il convient donc d'en définir les objectifs et d'y travailler avec autant de soin que sur les autres.

Je demanderais d'ailleurs bien à Bernard de concocter un CD en travaillant avec Stéphane Laurent sur des tests mettant en exergue dynamique, compression, transitoires et autres particularités de modulation pour travailler sur ces fameux programmes musicales (je mets un S car il faudrait des échantillons de styles différents). En dehors de cela l'auteur nous a montré qu'il fallait comprendre les protocoles qui permettent aux périphériques de se mettre en relation avec les aides auditives. Ce travail de mise en relation est souvent un peu fastidieux pour les gens qui ne sont pas nés avec le téléphone portable et j'en fais partie. J'étais donc très content d'entendre des explications allant dans ce sens.



Excellente cuvée et belle implication des auteurs dans leurs conférences respectives chacun avec sa personnalité. Donc en attendant le magnifique travail d'Yves Lasry sur les tests (travail qui sera accessible au mois de mars 2010, il faudrait demander au Président du Collège de faire concevoir un CD d'analyse des tests et à Bernard Hugon et à Stéphane Laurent et ainsi qu'à Léon Dodelé de nous préparer et de nous calibrer un CD pour paramétrer les programmes musique et parole dans la musique, de quoi s'occuper jusqu'au prochain EPU.

F. Degove

2

## EPU 2010

Comme chaque année, le Collège National d'Audioprothèse met en place un cycle de formation post-universitaire avec le concours des Directeurs des enseignements préparatoires au diplôme d'Etat d'Audioprothésiste.

Cette manifestation aura pour thème : **ACOUPHENES - HYPERACOUSIE** et se tiendra les vendredi 26 et Samedi 27 Novembre 2010 à la Cité des Sciences et de l'Industrie, Centre des Congrès de LA VILLETTE, 30, avenue Corentin Cariou à PARIS (19<sup>ème</sup>).

Ces journées seront rehaussées par une exposition des industriels fabricants et importateurs de matériels d'audioprothèse et d'audiophonologie.

**Renseignements :**

Danièle KORBA  
Collège National d'Audioprothèse  
10, rue Molière - 62220 CARVIN  
Tél. : 03 21 77 91 24  
Fax : 03 21 77 86 57  
E-mail : [College.Nat.Audio@orange.fr](mailto:College.Nat.Audio@orange.fr)  
[www.college-nat-audio.fr](http://www.college-nat-audio.fr)

3

**Congrès des audioprothésistes  
13, 14 et 15 mars 2010**
**CNIT - Paris La Défense****Samedi 13 mars 2010**

Conférence : « De l'aide auditive au cerveau ! Que nous offre la technique pour améliorer le service audioprothétique rendu ? »

*Paul AVAN, Président de la Journée Scientifique du Congrès - Laboratoire de Biophysique Sensorielle - Faculté de Médecine - Université d'Auvergne - Clermont-Ferrand*

**Dimanche 14 mars 2010 - Matin**

Remise du prix du Collège National d'Audioprothèse  
*Éric BIZAGUET, Président du Collège National d'Audioprothèse*

**Dimanche 14 mars 2010 - Après-midi**

« Mieux connaître, mieux comprendre pour mieux recevoir ses patients »  
*Sophie SCHMITT, Experte auprès de la*

*commission européenne pour les questions relatives aux nouvelles technologies appliquées aux seniors*

**Lundi 15 mars 2010**

Table ronde professionnelle  
*Benoît ROY, Président de l'UNSAF*  
De nombreux exposant, samedi et dimanche de 9h00 à 19h00 - Lundi de 9h00 à 17h00

**Ateliers**
**Dimanche 14 et lundi 15 mars 2010  
10h00-12h00 / 15h00-17h00**

La mesure in-vivo d'efficacité de l'appareillage auditif  
*Xavier DELERCE, Audioprothésiste D.E. Mont de Marsan*

**Dimanche 14 Mars 2010****10h00-12h00 / 15h00-17h00****Lundi 15 Mars 2010****10h00-12h00**

Bilan d'orientation prothétique du très jeune enfant  
*Thierry RENGLET, Audioprothésiste, Correspondant CNA, Belgique*

**Dimanche 14 Mars 2010****Lundi 15 Mars 2010****10h00-12h00 / 15h00-17h00**

Actualités en thérapie acoustique de l'acouphène  
*Philippe LURQUIN, Audiologiste - Chargé de cours, C.H.U. St-Pierre - Bruxelles*  
*Docteur DEBATY, ORL, C.H.U. St-Pierre - Bruxelles*

**Informations & Contacts**

**Retrouvez toutes les informations sur le congrès 2010 ainsi que le formulaire d'inscription sur notre site :**

[www.unsaf.org](http://www.unsaf.org)  
UNSAF : 66, rue Bernard Palissy - 37000 TOURS

E-mail : [contact@unsaf.org](mailto:contact@unsaf.org)  
Christine DAGAIN  
Chargée de l'organisation du Congrès des Audioprothésistes Français  
[christine.dagain@wanadoo.fr](mailto:christine.dagain@wanadoo.fr)

GROUPE SPAT  
Fanny HUBERT  
Responsable du Congrès  
34, rue de l'Église - 75015 PARIS  
Tél. : +33 1 44 26 26 27  
Fax : +33 1 45 54 23 86  
e-mail : [fhubert@spat.fr](mailto:fhubert@spat.fr)  
[www.spat.fr](http://www.spat.fr)

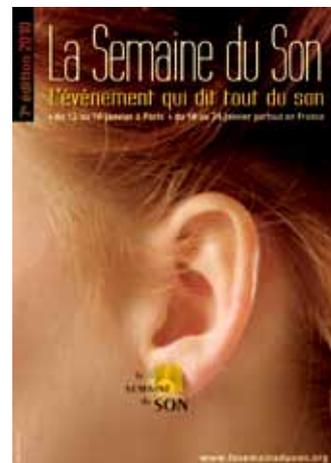
GROUPE SPAT  
Biba AGGAR

Responsable de l'Exposition  
34, rue de l'Église - 75015 PARIS  
Tél. : +33 1 44 26 26 25  
Fax : +33 1 45 54 23 86  
e-mail : [biba@spat.fr](mailto:biba@spat.fr)  
[www.spat.fr](http://www.spat.fr)

4

**Siemens partenaire  
de la Semaine du Son**

Comme en 2009, Siemens Audiologie sera partenaire en 2010 de la Semaine du Son.



A cette occasion, un kit de communication sera mis à disposition des audioprothésistes pour leur permettre d'afficher leur soutien.

Nous nous associerons, à Paris et en Province, aux nombreuses actions d'information, de prévention et de sensibilisation en disposant notre borne auditive dans les endroits à fort trafic.

**Renseignements :**

Nadège OUIDRANE  
Siemens Audiologie S.A.S.  
Marketing & Communication  
175 Bd Anatole France - B.P. 40  
93201 SAINT DENIS Cedex - France  
Tél. + 33 1 49 33 25 96  
Fax + 33 1 49 33 15 09



« Votre audioprothésiste,  
partenaire de la bonne audition.  
Découvrez les nouvelles solutions  
auditives numériques Siemens. »

SIEMENS



Mieux entendre tout en préservant  
l'environnement, c'est possible ?

Avec le chargeur  
Siemens, les aides  
auditives sont  
rechargées en  
quelques heures.  
Plus de pile à  
jeter, à acheter  
ni à remplacer !



Technologie  
rechargeable !

Offre de  
communication  
au dos...

## Chargeur Siemens. Pratique, économique, écologique.

Les français consomment en moyenne 40 millions de piles auditives chaque année. Siemens, N°1 mondial de la correction auditive, fabrique et commercialise des aides auditives rechargeables pour diminuer la consommation de piles et participer ainsi à la protection de l'environnement. Avec le chargeur Siemens, les aides auditives fonctionnent sur accumulateur. Il suffit de mettre les aides auditives dans le chargeur chaque soir. Elles sont éteintes automatiquement. Après seulement quelques heures de charge, elles sont prêtes à fonctionner toute la journée. Vos patients n'ont plus à se soucier des piles vides, à remplacer au cours d'un dîner ou d'une réunion. Ils profitent de la liberté d'une audition sans contrainte. La technologie rechargeable est une réponse aux attentes exprimées par les utilisateurs. Le chargeur fonctionne avec les aides auditives Siemens de dernière génération : Siemens Pure, Siemens Motion, Siemens Active.  
[www.siemens.fr/audiologie](http://www.siemens.fr/audiologie)

Answers for life\*

**SIEMENS**

# Contour S Series

Aussi beau à l'intérieur qu'à l'extérieur !

## Sweep™ Technologie

Première aide auditive dotée de la technologie tactile identique à celle utilisée par l'Ipod Touch ou l'iPhone d'Apple. Une surface tactile remplace les boutons ou molettes difficiles à utiliser. Le volume ou les programmes peuvent être ajustés par un simple contact du bout du doigt.

## Drive Architecture

Grâce à la Drive Architecture, les algorithmes fonctionnent en parallèle. Elle diminue significativement le temps de traitement des données tout en améliorant la consommation.

## AudioScape

Bénéficie de toutes les fonctions avancées pour apporter à vos patients un confort inégalé, une grande clarté sonore, une directionnalité améliorée et une meilleure perception vocale.

## Type de pile

Alimenté par une pile 13 qui prolonge son autonomie.



C'est plus intelligent. C'est plus simple. C'est Starkey.

 **series**  
DRIVE ARCHITECTURE

Experience more.™  **Starkey.**