ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

THÈSE PAR ARTICLES PRÉSENTÉE À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DU
DOCTORAT EN GÉNIE
Ph.D.

PAR
Cécile LE COCQ

COMMUNICATION DANS LE BRUIT:
PERCEPTION DE SA PROPRE VOIX ET REHAUSSEMENT DE LA PAROLE

MONTRÉAL, le 11 JANVIER 2010

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Frédéric Laville, directeur de thèse
Département de Génie Mécanique, École de technologie supérieure

M. Christian Gargour, codirecteur
Département de Génie Électrique, École de technologie supérieure

M. Marcel Gabrea, président du jury
Département de Génie Électrique, École de technologie supérieure

M. Philippe Guillemain, chargé de recherche, examinateur externe
Laboratoire de Mécanique et d'Acoustique, CNRS France

Mme Chantal Laroche, Ph.D., examinatrice externe
Programme d'Audiologie et d'Orthophonie, Université d'Ottawa

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 14 DÉCEMBRE 2009

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

AVANT-PROPOS

Les travailleurs de l'industrie œuvrent dans des milieux bruyant où ils risquent d'encourir des pertes auditives sévères. Au Québec environs 400 000 travailleurs (Girard et al., 2007) sont exposés à de tels risques. Afin de protéger l'audition des travailleurs, l'OMS (Organisation mondiale de la santé, WHO - World Health Organization) a émis des recommandations pour notamment réduire la dose de bruit à laquelle sont soumis les travailleurs : sur une journée, le niveau équivalent de bruit pendant 8 heures ne doit pas dépassé 85 dB(A) (WHO, 2001, Ch 4.). Le port de protecteurs auditifs est une solution simple qui permet, en choisissant adéquatement le type de protection auditive (bouchons d'oreilles, coquilles ou double protection), de respecter cette recommandation.

Aujourd'hui, on trouve sur le commerce un grand nombre de systèmes de protection auditive différents. Or, le plus souvent, ces protections sont inconfortables, et de plus, ce qui est dangereux, elles perturbent la communication entre les ouvriers ainsi que la perception des signaux d'alarme (Berger et al., 2000; Suter, 1992). C'est ainsi que beaucoup de travailleurs ne les portent pas, ou très peu. Les travailleurs ne sont donc pas protégés.

”Les meilleurs bouchons d'oreilles sont ceux que le travailleur de l'industrie portera” (NIOSH, 1998). Pour résoudre ce problème, la compagnie SONOMAX a initié le développement d'un bouchon d'oreille confortable et ”intelligent” : il permettra aux travailleurs de continuer à percevoir les signaux d'information utile, tels que la parole ou les signaux d'alarme, tout en protégeant leurs oreilles des forts niveaux de bruit. Aujourd'hui, les dispositifs de protection des oreilles qui sont sur le marché ne permettent pas toujours, tout en protégeant l'audition, de percevoir les signaux d'information utile. En effet, chez les sujets qui ont une audition “normale”, il a été démontré (Berger et al., 2000; Suter, 1992) que la reconnaissance de la parole ainsi que la perception des signaux avertisseurs ne sont en général pas perturbées. Néanmoins ces sujets ne trouve pas toujours la conversation facile et confortable et retirent leurs protecteurs pour communiquer. Par contre, pour les sujets qui possèdent une perte auditive, la problématique est légèrement différente vu qu'en général, le port de protecteurs auditifs peut perturber la perception de la parole et des signaux d'information utile. Le développement de bouchons d'oreilles ”intelligents” constitueraient une grande innovation dans le domaine de la protection auditive : les sujets qui ont une audition “normale” pourraient converser confortablement, et ceux qui possèdent une perte auditive pourraient percevoir les signaux d'information utile.

La compagnie SONOMAX a contacté l'ÉTS pour initier un partenariat de recherche. Le projet de recherche a alors été financé par SONOMAX et deux organismes subventionnaires. Le premier organisme subventionnaire a été l'IRSST qui a octroyé une bourse doctorale à Jérémie Voix. Le deuxième organisme subventionnaire a été le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) qui a octroyé des fonds à l'ÉTS pour un projet de Recherche et Développement Coopérative (RDC) en partenariat avec SONOMAX.

La première phase du projet de recherche de SONOMAX à l'ÉTS sur le développement de nouveaux bouchons d'oreilles confortables et ”intelligents” a débuté en 2000 avec le travail de thèse de Jérémie Voix (Voix, 2006) et portait sur la mise au point de l'embase auriculaire, l'élaboration d'un protocole objectif de mesure et de certification de l'atténuation effective procurée par un protecteur auditif intra-auriculaire et la prédiction de l'atténuation apportée par un bouchon fait sur mesure et muni d'un filtre.

La deuxième phase du projet de recherche SONOMAX, qui est l'objet de cette thèse, traite de la communication dans le bruit et vise à étudier la perception de notre propre voix lors du port de protecteurs auditifs du type bouchon d'oreille et à proposer un traitement électronique pour améliorer la compréhension de la parole des autres personnes.

REMERCIEMENTS

Je tiens tout d’abord à remercier mes directeurs Messieurs Frédéric Laville et Christian Gargour pour leur soutien inconditionnel tout au long de ces dernières années. Je les remercie pour la confiance qu’ils m’ont accordée dans toutes les étapes de ma recherche.

Merci au CRSNG et à la compagnie Sonomax pour leur soutien financier.

Un immense merci à Jérémie Voix, ami et collègue qui a toujours été de très bons conseils dans les moments difficiles.

Un grand merci à toute l’équipe de la compagnie Sonomax : Jean-Nicolas Laperle, Michael C. Turcot, Jérémie Voix, Remon Makarious, Katia Chabot, Mylène Landry, et tous les autres ... pour la confiance et le soutien qu’ils m’ont accordés.

Un remerciement particulier à Jean-Marc Lina pour son expertise mathématique et sa bonne volonté.

Merci au personnel technique de l’ETS : Serge Plamondon, Patrick Sheridan et tous les autres, sans l’aide de qui je n’aurais pas pu réaliser mes expériences.

Merci à tous mes sujets et à Alexandra Cloutier qui m’a assistée dans mes expériences.

Merci à tous mes amis Myriam, Nathalie, Sébastien ... et à tous ceux qui, à un moment donné, ont joué un rôle dans ce doctorat.

Un immense merci à toute ma famille, mes parents, mon frère, mes grands-parents pour leur soutien et leur amour.

COMMUNICATION DANS LE BRUIT: PERCEPTION DE SA PROPRE VOIX ET REHAUSSEMENT DE LA PAROLE

Cécile LE COCQ

RÉSUMÉ

La communication dans le bruit est un problème de tous les jours pour les travailleurs qui œuvrent dans des environnements industriels bruyants. Un grand nombre de travailleurs se plaignent du fait que leurs protecteurs auditifs les empêchent de communiquer facilement avec leurs collègues. Ils ont alors tendance à retirer leurs protecteurs et mettent ainsi leur audition à risque. Ce problème de communication est en fait double : les protecteurs modifient à la fois la perception de la propre voix du porteur, ainsi que la compréhension de la parole des autres personnes. Cette double problématique est considérée dans le cadre de cette thèse.

La modification de la perception de la propre voix du porteur des protecteurs est en partie due à l'effet d'occlusion qui se produit lorsque le conduit auditif est occlus par un bouchon d'oreille. Cet effet d'occlusion se traduit essentiellement par une amélioration de la perception des sons de basses fréquences internes à l'être humain (bruits physiologiques), et par une modification de la perception de la propre voix de la personne. Dans le but de mieux comprendre ce phénomène, suite à une étude approfondie de ce qui se trouve déjà dans la littérature, une nouvelle méthode pour quantifier l'effet d'occlusion a été développée. Au lieu d'exciter la boite crânienne du sujet au moyen d'un pot vibrant ou de faire parler le sujet, comme il se fait classiquement dans la littérature, il a été décidé d'exciter la cavité buccale des sujets au moyen d'une onde sonore. L'expérience a été conçue de telle manière que l'onde sonore qui excite la cavité buccale n'excite pas l'oreille externe ou le reste du corps directement. La détermination des seuils auditifs en oreilles ouvertes et occluses a ainsi permis de quantifier un effet d'occlusion subjectif pour une onde sonore dans le conduit buccal. Ces résultats ainsi que les autres quantifications d'effet d'occlusion présentées dans la littérature ont permis de mieux comprendre le phénomène de l'effet d'occlusion et d'évaluer l'influence des différents chemins de transmission entre la source sonore et l'oreille interne.

La compréhension de la parole des autres personnes est altérée à la fois par le fort niveau sonore présent dans les environnements industriels bruyants et par l'atténuation du signal de parole due aux protecteurs auditifs. Une possibilité envisageable pour remédier à ce problème est de débruiter le signal de parole puis de le transmettre sous le protecteur auditif. De nombreuses techniques de débruitage existent et sont utilisées notamment pour débruiter la parole en télécommunication. Dans le cadre de cette thèse, le débruitage par seuillage d'ondelettes est considéré. Une première étude des techniques "classiques" de débruitage par ondelettes est réalisée afin d'évaluer leurs performances dans un environnement industriel bruyant. Ainsi les signaux de paroles testés sont altérés par des bruits industriels selon une large de gamme de rapports signal à bruit. Les signaux débruités sont évalués au moyen de quatre critères. Une importante base de données est ainsi obtenue et est analysée au moyen d'un algorithme

de sélection conçue spécifiquement pour cette tâche. Cette première étude a permis de mettre en évidence l'influence des différents paramètres du débruitage par ondelettes sur la qualité de celui-ci et ainsi de déterminer la méthode "classique" qui permet d'obtenir les meilleures performances en terme de qualité de débruitage. Cette première étude a également permis de donner des guides pour la conception d'une nouvelle loi de seuillage adaptée au débruitage de la parole par ondelettes dans un environnement industriel bruité. Cette nouvelle loi de seuillage est présentée et évaluée dans le cadre d'une deuxième étude. Ses performances se sont avérées supérieures à la méthode "classique" mise en évidence dans la première étude pour des signaux de parole dont le rapport signal à bruit est compris entre -10 dB et 15 dB.

**COMMUNICATION IN A NOISY ENVIRONMENT:
PERCEPTION OF ONE'S OWN VOICE AND SPEECH ENHANCEMENT**

Cécile LE COCQ

ABSTRACT

Workers in noisy industrial environments are often confronted to communication problems. Lost of workers complain about not being able to communicate easily with their coworkers when they wear hearing protectors. In consequence, they tend to remove their protectors, which expose them to the risk of hearing loss. In fact this communication problem is a double one: first the hearing protectors modify one's own voice perception; second they interfere with understanding speech from others. This double problem is examined in this thesis.

When wearing hearing protectors, the modification of one's own voice perception is partly due to the occlusion effect which is produced when an earplug is inserted in the ear canal. This occlusion effect has two main consequences: first the physiological noises in low frequencies are better perceived, second the perception of one's own voice is modified. In order to have a better understanding of this phenomenon, the literature results are analyzed systematically, and a new method to quantify the occlusion effect is developed. Instead of stimulating the skull with a bone vibrator or asking the subject to speak as is usually done in the literature, it has been decided to excite the buccal cavity with an acoustic wave. The experiment has been designed in such a way that the acoustic wave which excites the buccal cavity does not excite the external ear or the rest of the body directly. The measurement of the hearing threshold in open and occluded ear has been used to quantify the subjective occlusion effect for an acoustic wave in the buccal cavity. These experimental results as well as those reported in the literature have lead to a better understanding of the occlusion effect and an evaluation of the role of each internal path from the acoustic source to the internal ear.

The speech intelligibility from others is altered by both the high sound levels of noisy industrial environments and the speech signal attenuation due to hearing protectors. A possible solution to this problem is to denoise the speech signal and transmit it under the hearing protector. Lots of denoising techniques are available and are often used for denoising speech in telecommunication. In the framework of this thesis, denoising by wavelet thresholding is considered. A first study on "classical" wavelet denoising technics is conducted in order to evaluate their performance in noisy industrial environments. The tested speech signals are altered by industrial noises according to a wide range of signal to noise ratios. The speech denoised signals are evaluated with four criteria. A large database is obtained and analyzed with a selection algorithm which has been designed for this purpose. This first study has lead to the identification of the influence from the different parameters of the wavelet denoising method on its quality and has identified the "classical" method which has given the best performances in terms of denoising quality. This first study has also generated ideas for designing a new thresholding rule suitable for speech wavelet denoising in an industrial noisy environment. In a second study, this new

thresholding rule is presented and evaluated. Its performances are better than the “classical method found in the first study when the signal to noise ratio from the speech signal is between -10 dB and 15 dB.

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

LISTE DES FIGURES

Page

INTRODUCTION

La présente thèse est une thèse par articles qui traite de la communication dans le bruit. Plus précisément sont étudiées la perception de notre propre voix lors du port de protecteurs auditifs de type bouchons d'oreilles et la compréhension des signaux de parole des autres personnes.

Dans cette introduction, une première partie présente le contexte général de la recherche, une deuxième partie les objectifs de la thèse, une troisième partie la problématique et la méthodologie, une quatrième et dernière partie la structure de la thèse.

0.1 Contexte

La communication dans le bruit est un problème auquel est confronté tout travailleur qui œuvre dans un environnement industriel bruyant (Berger et al., 2000; Suter, 1992). Tout d'abord, le milieu sonore industriel et les différents éléments qui le constituent sont présentés. Puis, les mécanismes de l'audition en oreille ouverte et occluse sont décrits. Finalement, les compromis entre santé et sécurité auxquels tout travail en environnement industriel bruyant est confronté sont explicités.

0.1.1 Le milieu sonore industriel

Le milieu sonore industriel est habituellement constitué de trois types principaux de signaux sonores que sont les bruits industriels, les signaux d'alarme et la parole (Berger et al., 2000; Suter, 1992). Ils seront présentés ici successivement.

0.1.1.1 Les bruits industriels

À ce jour, il n'existe pas de description précise d'un bruit industriel. De manière générale, il s'agit des bruits qui sont générés par les différentes machines et outils qui sont utilisés dans le milieu industriel. En traitement du signal, il est courant de classer les signaux selon leur stationnarité (Flandrin, 1998). Deux catégories de bruits industriels peuvent donc être considérées :

1. Les bruits non-stationnaires : ce sont par définition les bruits dont le contenu spectral varie au cours du temps ; par exemple les coups de marteaux données sur une plaque de métal génèrent un bruit d'impact qui est non-stationnaire ;
2. Les bruits stationnaires : ce sont par définition les bruits dont le contenu spectral reste inchangé au cours du temps ; par exemple les transformateurs électriques génèrent un bruit stationnaire.

La figure 0.1 présente les spectres moyens en tiers d'octaves de deux bruits industriels stationnaires enregistrés dans une usine de voitures issus de la base de données Noisex (NOISEX , 1990).

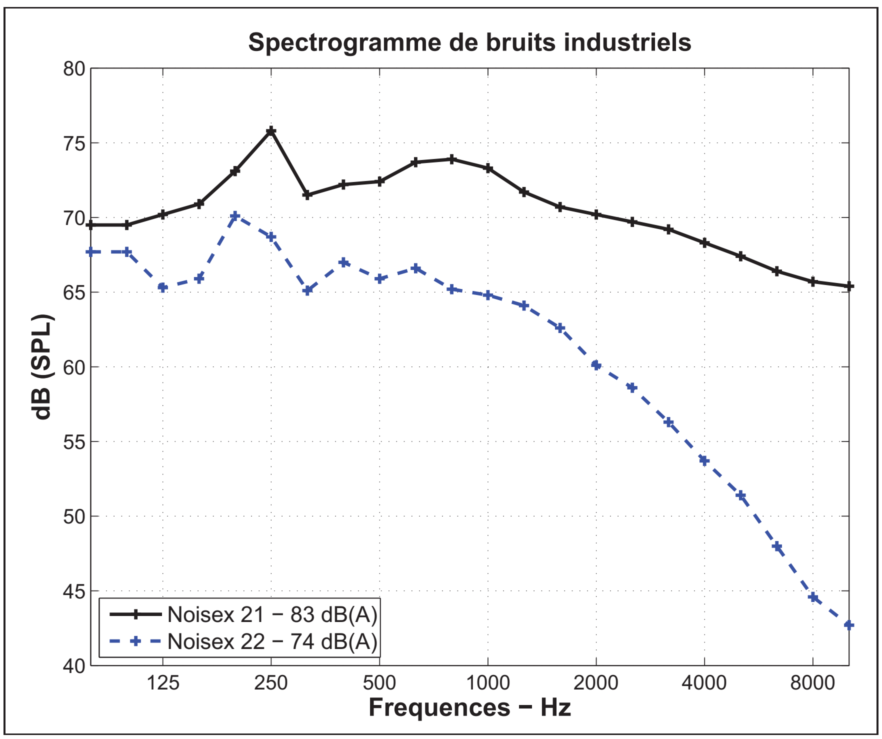

Figure 0.1 Spectres de deux bruits industriels enregistrés dans une usine de voitures, issus de la base de données Noisex (NOISEX , 1990).

Il est difficile de prédire le niveau sonore généré par une machine à l'oreille d'un travailleur. En effet pour ce faire il faudrait à la fois tenir compte de l'acoustique de la salle, de la localisation du travailleur et de la machine dans la salle, et du bruit généré par la machine. Tous les travailleurs d'une même entreprise ne seront donc pas soumis au même niveau sonore. Parmi la population à risque du canada (environs 2,200,000 travailleurs), d'après une étude réalisée par Voix et al. (2002) dans laquelle il avait regroupé des données statistiques issus de plusieurs sources et pour différents secteurs (construction, alimentation, imprimerie, textile, transports, et autres), 27.2% d'entre eux sont soumis a un niveau d'exposition supérieur à 100 dB (A).

0.1.1.2 Les signaux d'alarme

Dans le milieu industriel, de nombreux signaux d'alarme sont présents. Ils se distinguent selon leur rôle (Tran Quoc and Hétu, 1996) :

1. Un avertisseur de danger : pour prévenir les travailleurs d'un danger imminent. Il s'agit, par exemple, d'une alarme d'incendie ou d'un avertisseur de recul d'un véhicule.
2. Un signal indicatif impliquant une action : ces signaux permettent de prévenir le travailleur d'une action à réaliser. Il s'agit, par exemple, de la sonnerie d'un téléphone ou du signal sonore délivré par une machine pour indiquer un mauvais fonctionnement.

Les signaux d'alarme se distinguent également selon leur structure temporelle et fréquentielle :

1. Un signal continu : il peut se caractériser par un signal contenant plusieurs harmoniques qui seraient moduler en fréquences en fonction du temps. Certaines alarmes d'incendie ont cette caractéristique.
2. Un signal discontinu : il est constitué d'un court signal ou d'une série de courts signaux qui est répétée périodiquement en respectant un temps de pause entre chaque itération. L'avertisseur de recul d'un véhicule est ainsi défini.

Aujourd'hui, Il existe peu de normes de conception et d'implémentation de signaux d'alarme dans un milieu industriel. Selon l'entreprise considérée, un même signal d'alarme n'aura pas forcément la même signification. Par ailleurs, le nombre de signaux d'alarme différents présents dans un même milieu industriel est de plus en plus important. Or, en moyenne, un travailleur ne peut reconnaître facilement et rapidement que 7 signaux d'alarme différents (Tran Quoc and Hétu, 1996; Hétu, 1994).

0.1.1.3 La parole

Plusieurs niveaux de parole ont été définis (Webster, 1979) : le murmure, la voix normale, la voix élevée, la voix très forte, le cri et le niveau maximum de la voix. Ces niveaux ont été mesurés à 1 m devant le locuteur et sont présentés dans le tableau 0.1.

Tableau 0.1 Force de la voix associée aux différents niveaux de parole en dB (A) mesurés à 1 m devant le locuteur

Tiré de Webster (1979)

Force de la voix	**Niveau global - dB (A)**
Maximum	88
Cri	82
Très forte	74
Élevée	65
Normale	57
Détendue	50
Murmure	40

La figure 0.2 présente les spectres moyens de la parole à long terme pour différentes forces de la voix d'un homme mesurés à 1m devant le sujet. Il est à remarquer que l'allure du contenu spectral de la parole varie en fonction de la force de la voix employée par le locuteur (Hétu, 1994; Webster, 1979). La figure 0.3 nous donne la relation entre l'intelligibilité de la parole, la force de la voix et le rapport signal à bruit (Pickett, 1956) : quelque soit le rapport signal à bruit considéré, à très faible niveau, la voix est difficilement intelligible ; et quand le locuteur commence à forcer sa voix, l'intelligibilité de sa parole diminue.

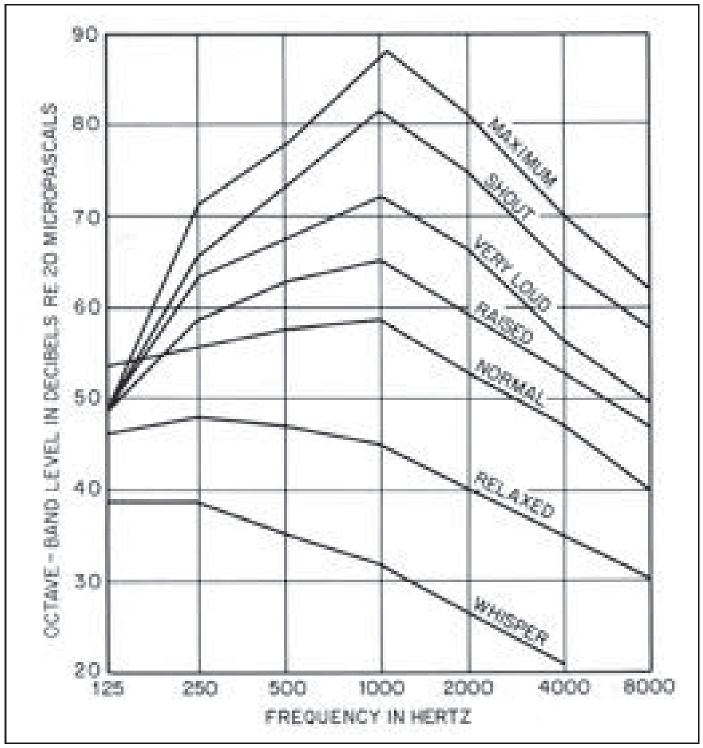

Figure 0.2 Spectres moyens de la parole à long terme pour différentes forces de la voix d'un homme mesurés à 1 m devant le sujet.
Tiré de Webster (1979)

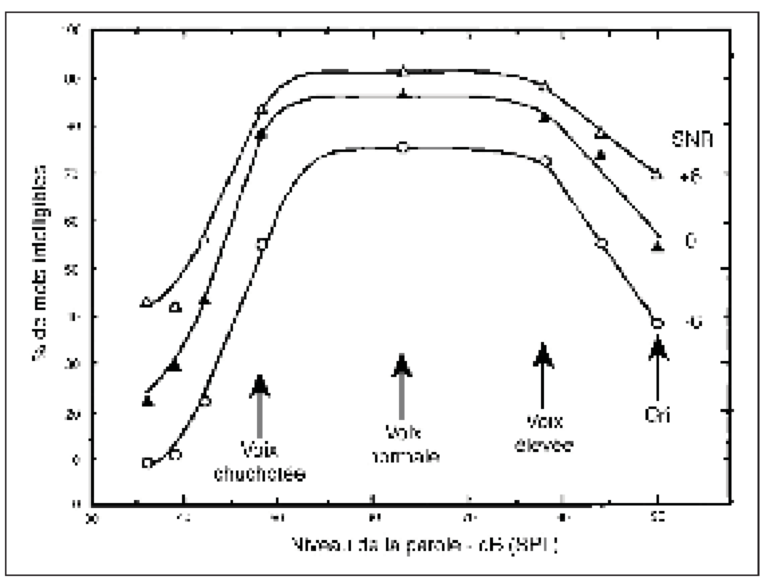

Figure 0.3 Relation entre l'intelligibilité de la parole, la force de la voix et le rapport signal à bruit (bruit blanc de 70 dB (SPL)).
Tiré de Pickett (1956)

0.1.2 Mécanisme de l'audition en oreille ouverte et occluse

Les mécanismes de l'audition en oreille ouverte et occluse sont ici brièvement présentés. Ils permettront par la suite de mettre en évidence des explications possibles de la diminution des capacités à percevoir, reconnaître et comprendre les signaux d'information utile, du type signaux d'alarme ou parole, lors du port de protecteurs auditifs (voir article #1). Sur la figure 0.4, sont tracés les différents chemins de transmission du son à l'oreille interne (transmission externe et interne) dans le cas d'une oreille ouverte ou occluse, sur le schéma anatomique du système auditif. Un signal qui arrive au récepteur auditif est la combinaison des signaux issus des différents chemins de transmission entre la source sonore et le récepteur auditif (Howell, 1985; Békésy, 1949, 1960a).

Dans le cas de la transmission externe en oreille ouverte, le parcourt le plus "classique" est celui appelé communément conduction "aérienne" : le son se propage dans le pavillon puis dans le conduit auditif externe pour être transmis à la cochlée par l'oreille moyenne. L'autre type de conduction est la conduction "osseuse" : un signal sonore est la plupart du temps au départ aérien, mais il peut à tout niveau du système auditif, mettre en vibration les os et se propager ainsi jusqu'à la cochlée.

Dans le cas de la transmission interne en oreille ouverte, transmission qui concerne tous les sons d'origine interne au corps humain (par exemple les bruits physiologiques, la voix), deux types de conduction sont mises en évidence (Békésy, 1960a; Tonndorf, 1972) : premièrement une conduction directe à la cochlée, et, deuxièmement, une conduction indirecte qui prend en compte le rayonnement dans le conduit auditif externe.

Dans le cas de la transmission externe en oreille occluse, c'est-à-dire quand un protecteur auditif de type bouchon d'oreille est introduit dans le conduit auditif externe, le chemin de transmission du son par la conduction "aérienne" est modifié pour la transmission externe et est remplacé par un ensemble de trois chemins de transmission (Berger, 1986) :

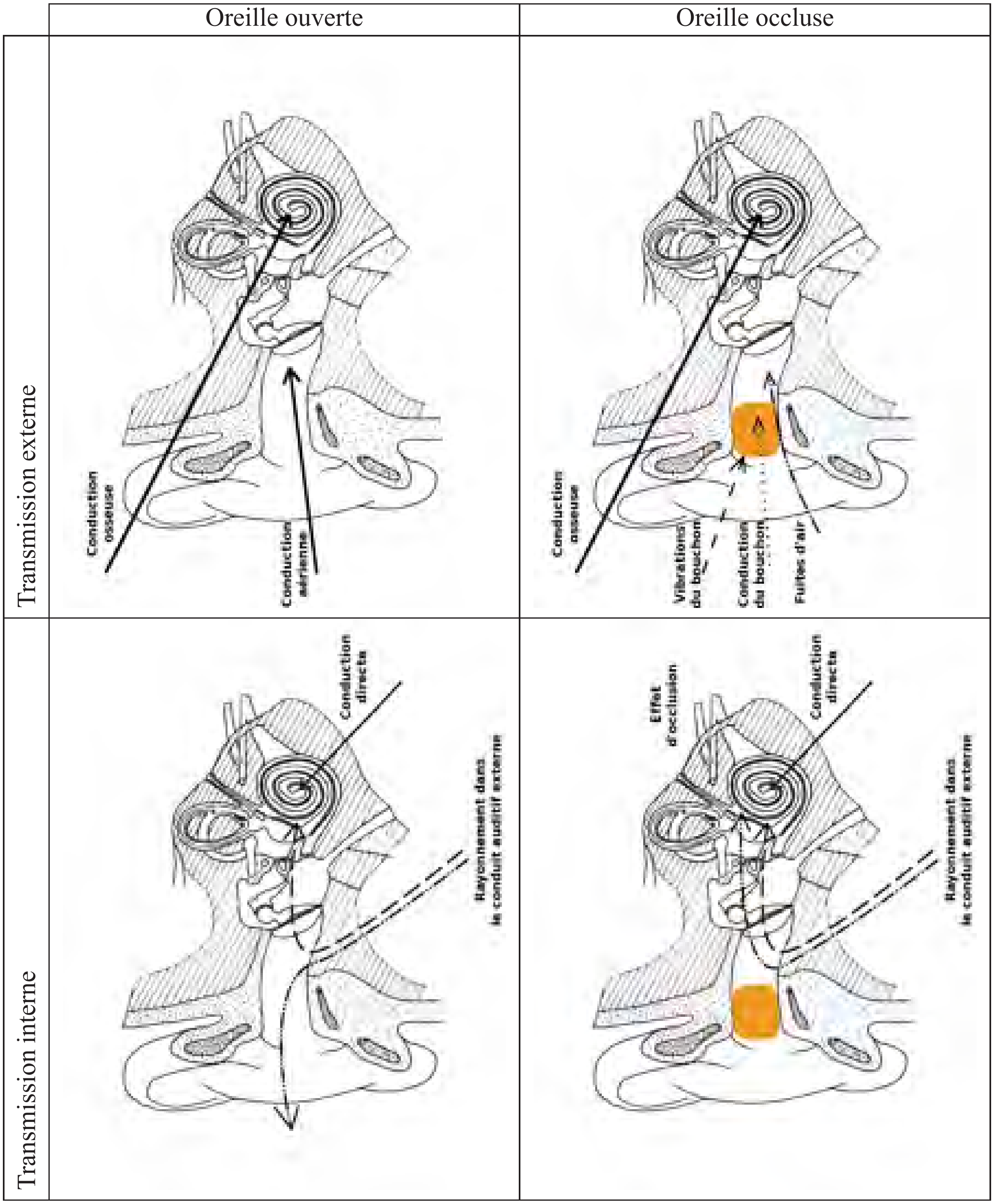

Figure 0.4 Représentation des différents chemins de transmission du son à l'oreille interne.

1. Un chemin de conduction par voie aérienne qui subsiste mais est réduit aux fuites d'air au niveau de l'étanchéité du protecteur auditif : l'onde acoustique se propage par voie aérienne entre le protecteur auditif et la paroi du conduit auditif, avant de se propager dans le conduit auditif externe restant ;
2. Un chemin de conduction au travers du protecteur auditif par voie solide : l'onde acoustique se propage par voie solide dans le matériau du protecteur auditif avant de se propager par voie aérienne dans le conduit auditif externe restant ;
3. Un chemin de conduction par vibration de l'ensemble de la structure du protecteur auditif : l'onde acoustique met en vibration l'ensemble du protecteur auditif par action sur sa paroi externe. La paroi interne est soumise au même mouvement vibratoire et elle va rayonner dans le conduit auditif externe restant.

Ces deux derniers chemins ont été distingués par Berger (1986), mais il s'agit en fait, d'un point de vue acoustique, du même type de transmission par voie solide par l'intermédiaire du bouchon. De plus, lors de cette transmission par voie solide par le bouchon, ce dernier va non seulement réémettre dans la partie aérienne du conduit auditif, mais aussi dans la peau et les structures sous-jacentes telles que cartilage et os auxquelles il est couplé. La conduction ”osseuse” s'en trouvera donc modifiée.

Dans le cas de la transmission interne, le signal qui rayonne dans le conduit auditif externe se dissipe en partie à l'extérieur en oreille ouverte. En oreille occluse, cette énergie est “piégée” à l'intérieur et est perçue par la personne sous la forme de l'effet d'occlusion (Stenfelt et al., 2003). L'effet d'occlusion se traduit en basses fréquences par une amélioration de la perception auditive et une augmentation du niveau sonore dans le conduit auditif (Stenfelt et al., 2003). Quand une personne parle, elle perçoit sa voix, à la fois par la propagation de sa voix à l'extérieur de la tête (transmission externe), à la fois par la voie interne (transmission interne). Lorsqu'une personne porte des protecteurs auditifs, elle ne perçoit pas sa voix de la même manière qu'en oreilles ouvertes pour deux raisons : premièrement, la propagation de sa voix par conduction “aérienne” se trouve atténuée par le protecteur auditif. Deuxièmement, l'effet d’occlusion (Stenfelt et al., 2003), mentionné précédemment, va augmenter la perception des basses fréquences de sa propre voix.

0.1.3 Compromis entre santé et sécurité

La partie 0.1.1 a permis d’exposer brièvement la complexité du contexte sonore en milieu industriel. Trois types de signaux que sont les bruits industriels, les signaux d’alarme et la parole sont présents. En raison notamment de leur niveau sonore et de leur présence souvent continue, les bruits industriels sont dominants. Le niveau sonore global en milieu industriel est donc égal au niveau sonore dû au bruit industriel seul. Ces forts niveaux sonores peuvent entraîner des pertes auditives chez les travailleurs qui sont présents dans ce contexte sonore pendant de longues périodes. La nécessité de protéger leur audition pour des raisons de santé sera présentée dans une première partie. Bien que ce soit les bruits industriels qui dominent les milieux sonores industriels, ce n’est pas leur perception qui est la plus importante pour la sécurité des travailleurs, mais la perception des signaux d’alarme et l’intelligibilité de la parole. Ces deux problématiques seront présentées dans la deuxième partie.

0.1.3.1 Protection de l’audition

L’audition est une caractéristique sensible de l’être humain qui a ses limites à la fois du côté des très faibles niveaux sonores que des très forts niveaux sonores. Sur la figure 0.5 est représenté les seuils d’audition minimum en champ diffus (MAF, minimal auditory field) et sous écouteurs (MAP, minimal auditory pressure) ainsi que les seuils d’inconfort et de douleur en niveau de pression exprimé en environnement industriel bruyantSPL).

Afin de protéger l’audition des travailleurs, des lois ont été proposées pour limiter la durée et le niveau d’exposition au bruit. Au delà de ces limites, ils doivent protéger leur audition au moyen de protecteurs auditifs. Il n’existe pas de législation internationale sur la protection auditive, et chaque pays choisit ses propres règlements. Le table 0.2 présente les règlements de quelques pays.

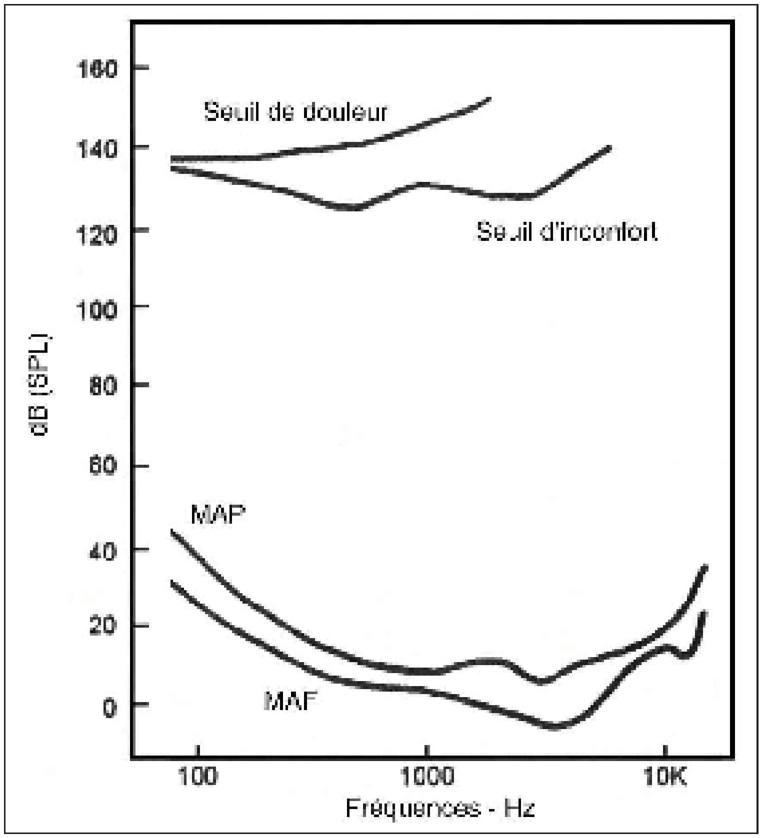

Figure 0.5 Seuil d'audition binaural minimum (MAF, minimal auditory field) d'après Robinson and Dadson (1956), seuil d'audition monaural minimum (MAP, minimal auditory pressure) d'après Dadson and King (1952), seuil d'inconfort d'après Wegel (1932) et seuil de douleur d'après Békésy (1960b).
Tiré de Gelfand (2004)

Tableau 0.2 Quelques législation sur la protection auditive (Johnson et al., 2001; Suter, 2000; Neitzel, 2009)

Origine de la législation	**Niveau moyen d'exposition pendant 8h en environnement industriel bruyantA)**
OMS (Johnson et al., 2001)	85
Canada (Fédéral)	87
Canada (Ontario, Québec, Nouveau-Brunswick)	90
USA	90
France	85

Selon l'Organisation Mondiale de la Santé (OMS), un niveau moyen d'exposition pendant 24 heures de 70 environnement industriel bruyantA) n'entraîne pas de perte auditive pour la majeure partie de la population (WHO, 1999), ce qui équivaut à un niveau de 75 dB (A) pendant 8 heures, si les 16 heures restantes sont à un niveau négligeable. L'OMS recommande qu'au delà d'un niveau moyen d'exposition de 85 environnement industriel bruyantA) pendant 8 heures (Johnson et al., 2001), les travailleurs protègent leur audition au moyen de protecteurs auditifs. Mais se protéger l'audition n'est pas tout, encore faut-il bien se protéger. Tous les protecteurs auditifs (bouchons d'oreille ou coquilles) n'apportent pas la même protection. Sur la figure 0.6 sont représentées les atténuations obtenues pour 4 différents protecteurs auditifs tirés du Compendium du NIOSH (2009) : un protecteur qui possède une faible atténuation, un protecteur qui possède une forte atténuation, et deux protecteurs qui possèdent une atténuation intermédiaire et qui dont l'allure de l'atténuation en fonction des fréquences diffère.

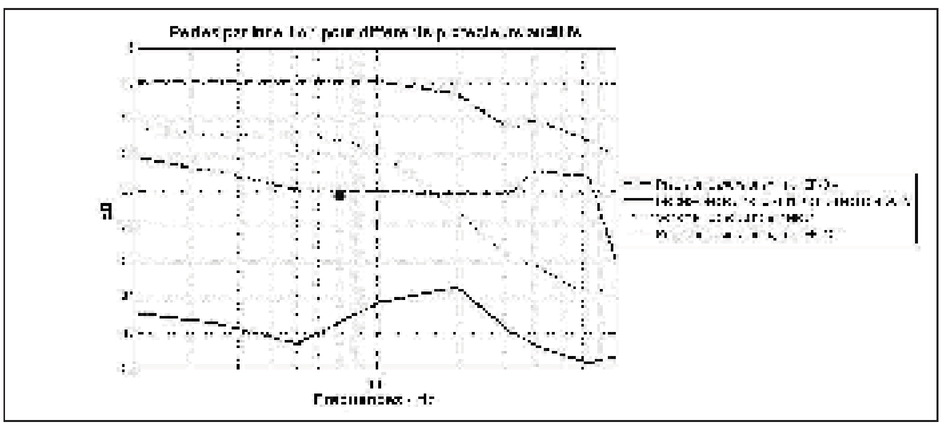

Figure 0.6 Atténuation obtenue pour différents protecteurs auditifs choisis dans le Compendium du NIOSH (2009) de manière à présenter l'ensemble des possibilités.

Certains protecteurs auditifs possèdent une atténuation (lors de mesures en laboratoire) qui peut atteindre 40 dB, alors que d'autres se limitent à 10 dB d'atténuation. Si la protection choisie n'est pas suffisante, le travailleur risque de subir des pertes auditives. Par contre à l'inverse, si la protection choisie est trop importante, le travailleur n'a en général rien à craindre pour son audition, mais peut se retrouver isolé du milieu sonore qui l'entoure. Il risque alors de ne plus entendre tous les signaux d'information utile, que ce soit le bruit d'une machine qui s'emballe, un signal de parole ou un signal d'alarme ; il met en danger sa sécurité. Le tableau 0.3 présente, d'après les recommandations EN458 (1993, 1996), et CSA (2002), le niveau de pression qui doit être présent sous le protecteur pour avoir une protection adéquate.

Tableau 0.3 Niveaux de protection tels que définis par les recommandations EN458 (1993, 1996), et CSA (2002)

Niveau de pression résiduel sous le protecteur environnement industriel bruyantA)	**Niveau de protection**
85 +	Insuffisant
80 - 85	Acceptable
75 - 80	Optimal ou Idéal
70 - 75	Acceptable
Inférieur à 70	Surprotection

0.1.3.2 Perception des signaux d'information utile

Un signal d'information utile sera toujours mieux perçu dans le silence que dans le bruit. L'influence du bruit ainsi que du port de protecteurs auditifs sur la perception des signaux d'information utile (signaux d'alarme et/ou parole) est présentée ici. Dans cette section, la notion de percevoir, reconnaître et comprendre les signaux d'information utile est utilisée. Ces trois termes indiquent des niveaux graduels d'interprétation des signaux. Percevoir correspond à entendre quelque-chose sans savoir de quoi il s'agit. Reconnaître consiste à identifier un son entendu comme étant un signal de parole d'un homme par exemple, mais sans comprendre ce qui est dit. Finalement, comprendre implique à la fois que le sujet a tout d'abord perçu et reconnu le signal sonore, mais que, en plus, il a pu assimiler toute la signification du message.

Saturation du système auditif

Il est à noter que le système auditif humain, comme tout système d'acquisition de signal sonore, peut saturer : si le niveau sonore global est trop élevé, la cochlée ne sera plus capable de traiter correctement les informations reçues, les capacités à percevoir et surtout à comprendre et reconnaître les signaux d'information utile seront amoindries (Robinson and Casali, 2000).

Le port de protecteurs auditifs permet, en général, de diminuer le niveau sonore global dans l'oreille et supprime ainsi la saturation de la cochlée. Si l'atténuation des protecteurs auditifs était constante pour toutes les bandes d'octave, ceux-ci permettraient d'améliorer les capacités à percevoir, reconnaître et comprendre les signaux d'information utile. Comme nous le voyons sur la figure 0.6, l'atténuation des protecteurs auditifs n'est pas constante pour l'ensemble des fréquences, mais à tendance, au contraire, à atténuer plus les hautes fréquences que les basses fréquences. Ceci pourrait diminuer les capacités à percevoir, reconnaître et comprendre les signaux d'information utile, notamment chez les personnes qui affichent des pertes auditives (Robinson and Casali, 2000). En effet les hautes fréquences sont très importantes du point de vue de l'intelligibilité de la parole.

Masquage des signaux d'information utile

Quand deux signaux sonores sont présents simultanément dans un environnement, ils auront tendance à se masquer. Le signal d'information utile est appelé le signal masqué, le bruit est appelé le signal masquant. La figure 0.7 présente l'allure du masquage produit par différents tons purs et pour différents niveau du signal masquant.

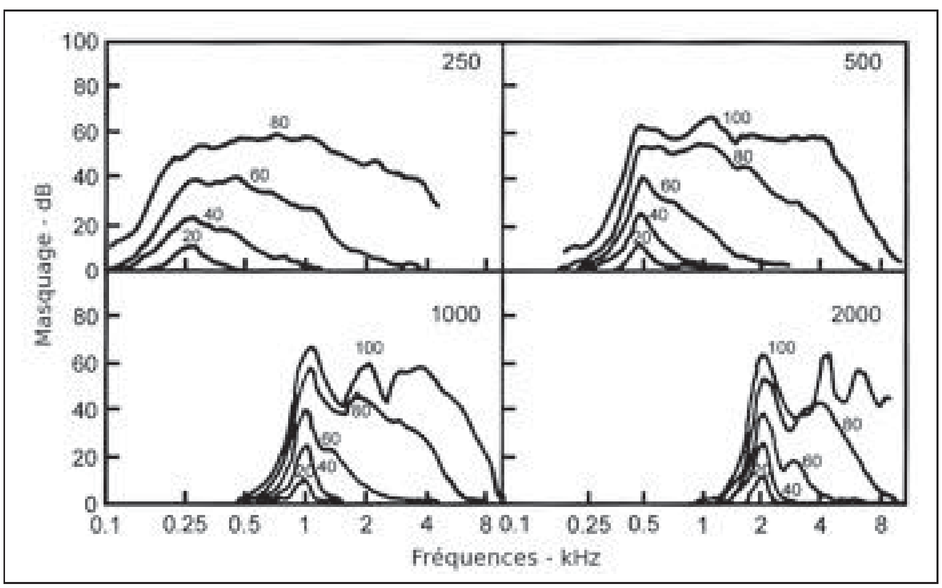

Figure 0.7 Allure du masquage produit par différents tons purs et pour différents niveaux du signal masquant.
Tiré de Gelfand (2004)

Le seuil de perception du signal masqué augmente linéairement en fonction du niveau sonore du signal masquant lorsque les deux signaux sont centrés sur la même fréquence (Gelfand, 2004; Robinson and Casali, 2000). Si on considère un bruit à bande étroite comme signal masquant, la plage fréquentielle de masquage sera d'autant plus importante que le niveau sonore du signal masquant sera élevé (Gelfand, 2004; Robinson and Casali, 2000). De plus, il est à remarquer que la plage fréquentielle de masquage n'est symétrique que pour un faible niveau sonore du signal masquant et devient très rapidement asymétrique : pour les fréquences inférieures, la plage fréquentielle est très étroite et chute très rapidement ; pour les fréquences supérieures, la plage de masquage est d'autant plus large que le niveau sonore du bruit masquant est élevé (Gelfand, 2004; Robinson and Casali, 2000). Ainsi un bruit à bande étroite centrée sur 200Hz générera à fort niveau (de l'ordre de 80 à 100 environnement industriel bruyantA)) la même plage fréquenctielle de masquage qu'un bruit large bande de niveau équivalent en énergie (Gelfand, 2004; Robinson and Casali, 2000).

Les bruits industriels possèdent en général une forte énergie en basses fréquences. Sur la figure 0.1, il est à remarquer, pour les deux bruits industriels présentés, que la partie la plus énergétique du spectre se situe en dessous de 1000Hz, avec un pic d'énergie dans les bandes de tiers d'octave centrés sur 250 et 200Hz. L'effet de masquage provoqué par ces bruits couvrira donc presque la totalité du spectre audible et altérera les capacités à percevoir, reconnaître et comprendre les signaux d'information utile. Le port de protecteurs auditifs permet de diminuer cet effet de masquage, vu que le niveau du signal masquant sera diminué. Toutefois, les protecteurs auditifs ont en général une atténuation non constante qui est plus importante pour les hautes fréquences (voir figure 0.6). La diminution de l'effet de masquage apporté par le port de protecteurs auditifs est alors moins bénéfique pour le travailleur, surtout si celui-ci est atteint d'une perte auditive (Tran Quoc and Hétu, 1996; Suter, 1992).

Perception des signaux d'alarme

Comme il a été présenté précédemment, la saturation de la cochlée due à un fort niveau sonore et l'effet de masquage dû au bruit industriel élevé entraînent une diminution de la capacité à percevoir, reconnaître et comprendre les signaux d'alarme. Le port de protecteurs auditifs n'améliorera pas forcément ces capacités, surtout chez des travailleurs qui sont atteints d'une perte auditive (Tran Quoc and Hétu, 1996; Suter, 1992). Par ailleurs, la conception des signaux d'alarme et de leur niveau sonore est laissée au hasard et n'est pas réalisée en fonction du contexte sonore la plupart du temps (Tran Quoc and Hétu, 1996; Hétu, 1994). Un niveau sonore adéquat pour un signal d'alarme doit être compris entre $+10$ dB et $+25$ dB au dessus du seuil de masquage pour un homme ayant une audition normale (Hétu, 1994) ou entre $+13$ dB et $+25$ dB au dessus du bruit avec un niveau maximal absolu de 105 environnement industriel bruyantSPL) (Tran Quoc and Hétu, 1996). Une étude des signaux d'alarme d'une aciérie a permis de mettre en évidence qu'uniquement 50% des signaux d'alarme avaient un niveau sonore adéquat ; 15% étaient trop faible pour être perçus, 25% étaient trop fort et pouvaient entraîner une gêne et de la fatigue auditive (Hétu, 1994). Ce problème de sécurité associé à la perception des signaux d'alarme ne peut pas être entièrement résolu par des protecteurs auditifs "intelligents" qui détecteraient les signaux d'alarme et les transmettraient au travailleur. En effet, un signal d'alarme qui est noyé dans le milieu sonore ne sera pas mieux détecté par un système de microphones que par l'oreille humaine. Tran Quoc and Hétu (1996) ont proposé une série de règles pour aider à concevoir les signaux d'alarme en fonction du milieu industriel considéré et ainsi garantir que ces signaux d'alarme seront entendus par les travailleurs, qu'ils possèdent une perte auditive ou non. Dans l'état actuel des choses, les signaux d'alarme ne respecte pas en général ces règles de conception. Les ouvriers peuvent donc rencontrer des difficultés à percevoir ces signaux, qu'ils portent leur protecteur auditif ou non. Même un système électronique muni de microphone pourrait avoir du mal à les détecter, en raison de leur trop faible niveau ou de leurs caractéristiques fréquentielles pas assez différenciées par rapport au bruit industriel environnant.

Intelligibilité de la parole

L'intelligibilité de la parole est ici considérée selon deux points de vue : premièrement la perception de la parole d'autrui, deuxièmement la perception ne notre propre voix.

Pour la perception de la parole d'autrui, de même que pour les signaux d'alarme, la saturation de la cochlée due à un fort niveau sonore et l'effet de masquage dû au bruit industriel élevé entraînent une diminution de la perception et de l'intelligibilité de la parole qui ne seront pas forcément améliorées par le port de protecteurs auditifs (Tran Quoc and Hétu, 1996; Suter, 1992). Par ailleurs, en milieu bruyant, une personne qui ne porte pas de protecteurs a tendance à élever la voix pour se faire entendre. Si elle doit crier, l'intelligibilité de ses paroles diminue (Hétu, 1994; Pickett, 1956).

Du point de vue de la perception de notre propre voix, lorsqu'un travailleur porte des protecteurs auditifs, sa perception de sa propre voix est modifiée (cf. partie 0.1.2), il aura alors tendance à modifier le contenu spectral de sa voix pour retrouver sa voix "normale", ce qui peut également diminuer l'intelligibilité de ses paroles pour un autre auditeur. De plus, l'effet d'occlusion donne l'impression au porteur des bouchons qu'il parle plus fort qu'il ne parle réellement ; il a alors tendance à diminuer le niveau global de sa voix, ce qui n'aide pas à l'intelligibilité de ses paroles pour un autre auditeur. Le rapport signal à bruit de sa voix par rapport au bruit ambiant est diminué, ce qui diminue d'autant l'intelligibilité de ses paroles pour un autre auditeur (Gelfand, 2004; Robinson and Casali, 2000).

0.1.3.3 Synthèse : Compromis entre santé et sécurité

La figure 0.8 présente un diagramme récapitulatif des compromis entre santé et sécurité.

À un fort niveau sonore, le port de protecteurs auditifs est indispensable pour protéger l'audition des travailleurs ; il faut faire attention d'avoir une protection adéquate à la fois pour protéger suffisamment l'audition, sans tomber dans un excès de protection qui entraîne des risques pour la sécurité du travailleur. Toutefois, dans certains cas, notamment chez les travailleurs ayant une perte auditive, le port de protecteurs auditifs, même bien choisis, peut entraîner une diminution des capacités à percevoir, reconnaître et comprendre les signaux d'information utile ; la sécurité du travailleur est alors compromise. Pour la perception des signaux d'alarme, une partie des difficultés de reconnaissance est due à une mauvaise conception de ces derniers qui se retrouvent noyés dans le bruit. En ce qui concerne l'intelligibilité de la parole,

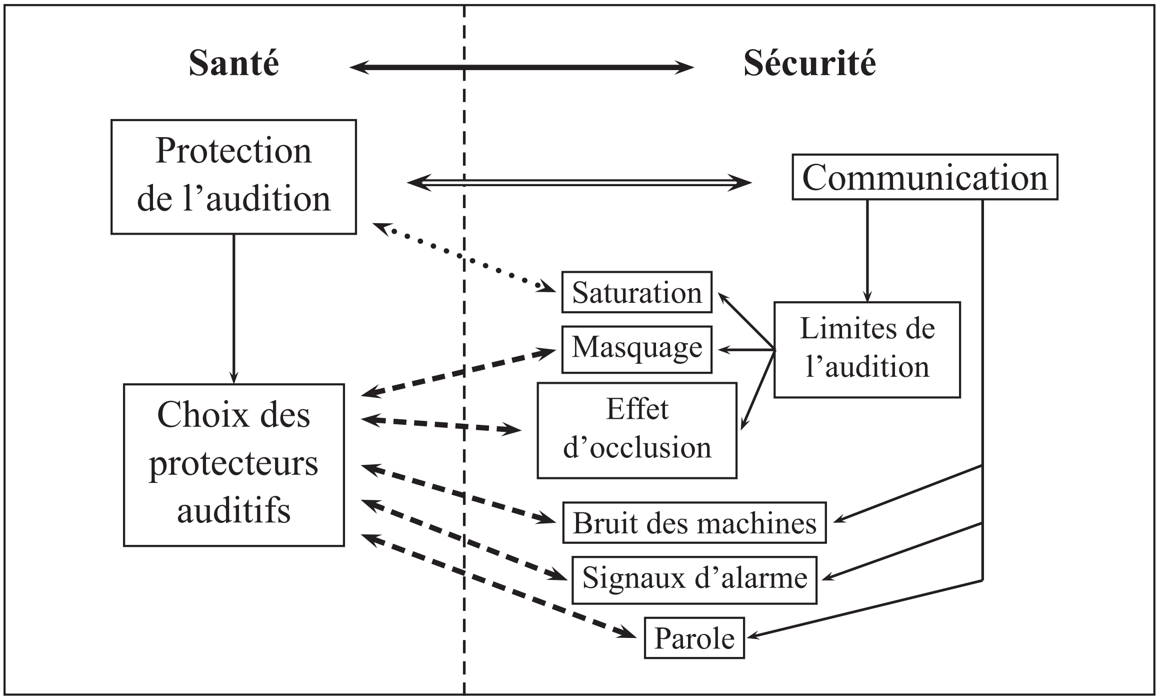

Figure 0.8 Diagramme récapitulatif des compromis entre santé et sécurité

les sources des difficultés de compréhension se situent à la fois du côté du locuteur qui modifie le spectre de sa voix et du côté de l'auditeur qui porte des protecteurs auditifs.

0.2 Objectif

L'objectif principal de cette thèse est d'améliorer la communication verbale des travailleurs dans le bruit en milieu sonore industriel, à la fois du point de vue du locuteur et de l'auditeur. Pour le locuteur, la perception de sa propre voix ne doit pas être modifiée ou très peu par le port de protecteur auditifs ; ainsi il pourra parler normalement et ses paroles seront très intelligibles. Pour l'auditeur, il s'agit d'améliorer l'intelligibilité de la parole d'autrui en milieu sonore industriel ; pour se faire il s'agit de débruiter la parole du locuteur et de la réémettre sous les protecteurs auditifs que porte l'auditeur. Deux sous-objectifs se dégage ainsi de l'objectif principal.

1. Pour le locuteur, afin d'améliorer la perception de sa voix lors du port de protecteurs auditifs, une étude de l'effet d'occlusion est réalisée pour comprendre les altérations qu'il entraine sur la perception de notre propre voix.
2. Pour l'auditeur, afin d'améliorer l'intelligibilité de la parole d'autrui, le débruitage de la parole en milieu sonore industriel est étudié et évalué. Le cas du débruitage de la parole par seuillage d'ondelettes est ici considéré.

Le premier sous-objectif traite d'aspects physiologiques de l'audition étudiés de manière expérimentale et fondamentale. Le deuxième sous-objectif consiste en une recherche appliquée et théorique en traitement du signal par ondelettes. Le premier sous-objectif est traité dans le cadre du premier article de la thèse. Le deuxième sous-objectif est étudié dans les articles #2 et #3 de la thèse.

0.3 Problématique et Méthodologie

Dans cette section, la problématique et la méthodologie suivies tout au cours de la thèse pour chacune des deux parties sont ici présentées et expliquées.

0.3.1 Effet d'occlusion

La question pour laquelle une réponse était recherchée est la suivante : Comment le port de protecteurs auditifs intra-auriculaires modifie la perception interne de notre propre voix ?

Tout d'abord il s'agissait de trouver dans la littérature des éléments de réponse. L'effet d'occlusion est apparu comme un des facteurs de modification de la perception interne de notre propre voix. Il a beaucoup été étudié dans le cas d'une excitation par conduction osseuse. Des quantifications subjectives et objectives de l'effet d'occlusion sont ainsi disponibles. Une quantification objective de l'effet d'occlusion présent quand un sujet parle a également été réalisée. Par contre, la littérature ne fournit pas de quantification subjective de l'effet d'occlusion pour la voix du sujet.

Afin de combler ce manque, une nouvelle technique pour quantifier l'effet d'occlusion a été conçue : le matériel expérimental a été développé de telle manière qu'une onde sonore émise dans la bouche du sujet ne puisse pas être transmise à la cochlée par la voix externe, mais qu'elle soit uniquement transmise par voix interne. Une quantification à la fois subjective et objective de l'effet d'occlusion obtenu ainsi avait été prévue et réalisée expérimentalement.

Pour ce qui est de la quantification objective, une fois que toute les données avaient été récupéré pour les différents sujets, les résultats se sont avérés faux au dessus de 1,000 Hz : exactement la même courbe était obtenue pour tous les sujets. Nous nous sommes aperçus à ce moment là qu'il s'agissait d'un bruit électrique qui était transmis sur les microphones lors de l'acquisition. Ces résultats expérimentaux étaient donc non utilisables car non fiables. La décision a alors été prise de compléter la thèse sans ces mesures afin de ne pas reporter le dépôt et la soutenance de celle-ci.

Les résultats de la quantification subjective ont été analysés. Une caractérisation de l'effet d'occlusion, de la perception interne de la voix et de l'influence des différents chemin de conduction a été entreprise à partir de ces résultats ainsi que des quantifications déjà présentes dans la littérature. L'article #1 présente le protocole expérimental, les résultats obtenus et les conclusions qui ont été tirées.

0.3.2 Débruitage de la parole en milieu industriel

De nombreuses méthodes de débruitage de la parole existent et sont principalement utilisées dans le domaine des télécommunications. Dans un environnement industriel bruyant, les contraintes sonores ne sont pas les mêmes, il est donc a priori difficile de prévoir les performances d'un algorithme de débruitage de la parole, conçu pour le domaine des télécommunications, dans un milieu industriel.

Parmi les méthodes de débruitage de la parole présentes dans le littérature, le débruitage par ondelette a été choisi pour plusieurs raisons. Tout d'abord, la transformée en ondelettes réalise une décomposition fréquentielle proche de celle de la cochlée. Par ailleurs les bruits industriels sont en général non blanc, et dans une même entreprise plusieurs bruits différents sont présents. L'ondelette a la capacité de s'adapter selon les déplacements du travailleur, et selon les composantes fréquentiels du bruit.

Par conséquent, Les techniques "classiques" de débruitage par ondelettes de la parole ont été testées et évaluées dans un environnement industriel bruyant. Au total, 1 296 méthodes de débruitage par ondelettes ont été testées sur 8 200 signaux de parole bruités. Vu l'importance de la base de données ainsi obtenue à analyser, il n'était plus envisageable de faire une étude subjective des performances. Quatre critères ont donc été utilisés pour quantifier les performances des résultats obtenus. Un algorithme de sélection spécialement conçu à cet effet a ensuite permis de faire une analyse objective pour mettre en évidence la ou les méthodes qui semblent les plus prometteuses pour débruiter un signal de parole en milieu industriel bruité.

Cette étude exploratoire a permis de mettre en évidence l'influence des différents paramètres du débruitage par ondelettes, et a permis, par la suite, de proposer et de concevoir une nouvelle règle de seuillage spécifiquement adaptée au milieu industriel bruité.

0.4 Structure de la thèse

La présente thèse est structuré comme suit. Le premier chapitre constitue l'article #1 intitulé "Subjective characterization of earplugs' occlusion effect using an external acoustical excitation of the mouth cavity" qui répond au premier sous-objectif de la thèse. Pour le deuxième sous-objectif de la thèse, deux articles permettent d'y répondre et sont présentés dans les chapitres 2 et 3. Le deuxième chapitre est l'article #2 intitulé "Wavelet speech enhancement for industrial noise environments". Cet article traite de l'étude exploratoire réalisée ainsi que l'algorithme de sélection conçu pour analyser les performances des résultats obtenus. Le troisième chapitre est l'article #3 intitulé "A wavelet speech thresholding rule for denoising in industrial environments". Il explicite la nouvelle loi de seuillage proposée et présente les performances qu'elle permet d'obtenir. Finalement, une conclusion générale de la thèse permet de faire le point sur le travail qui a été réalisé et les avenues à envisager dans l'avenir.

CHAPITRE 1

ARTICLE #1

"SUBJECTIVE CHARACTERIZATION OF EARPLUGS' OCCLUSION EFFECT USING AN EXTERNAL ACOUSTICAL EXCITATION OF THE MOUTH CAVITY"

Cécile Le Cocq, Frédéric Laville, Christian Gargour,

École de technologie supérieure, Université du Québec,

Montréal (Québec), Canada, H3C 1K3

This article has been submitted to the Journal of the Acoustical Society of America,

16 September 2009.

Résumé

L'occlusion du conduit auditif par le port d'une prothèse ou d'un protecteur auditif crée un inconfort chez les usagers qui est dû, entre autres, à une modification de la perception de sa propre voix. Ce phénomène s'appelle effet d'occlusion (OE - occlusion effect). Cet article propose une caractérisation de l'effet d'occlusion associé à la voix basée sur un modèle de transmission interne du son divisé en un chemin de conduction de la voix par le corps (VBC - voice body conduction) et en un chemin de conduction de la voix par l'air et le corps (VABC - voice air and body conduction). Une mesure subjective de ce dernier chemin est proposée: elle utilise un haut-parleur placé à l'entrée de la bouche. Les résultats issus de cette mesure et ceux rapportés dans la littérature sont utilisés afin de caractériser les chemins internes de conduction de la voix. En basses fréquences, lorsque les oreilles sont occluses, le chemin indirect domine le chemin direct et les OE objectif et subjectif sont positifs. En hautes fréquences, lorsque les oreilles sont non occluses, le chemin VBC direct domine le chemin VBC indirect, alors que lorsque les oreilles sont occluses, c'est le chemin VABC indirect qui domine le chemin VABC direct. À ces fréquences, les OEs objectifs relatifs aux chemins VBC et VABC et à la voix ainsi que l'OE subjectif associé au chemin VABC sont négatifs, tandis que l'OE subjectif relatif au chemin VBC est de l'ordre de zéro.

Abstract

The occlusion of the ear canal by hearing aids or hearing protectors results in an occlusion effect (OE) which creates a discomfort to their users due to the resulting changes in their own voice perception. In this paper, a characterization of the voice OE is proposed, based on a transmission scheme where the internal ear sound path is subdivided into the voice body conduction (VBC) path and the voice air and body conduction (VABC) path. A subjective measurement of this path is presented where a speaker is placed at the mouth entrance. The results from these measurements and others reported in the literature, are used to characterize the internal voice paths. At low frequencies, in an occluded ear, the indirect path is dominant over the direct one and the objective and subjective OE are positive. At high frequencies, in an open ear, the VBC direct path is dominant over the indirect one, while, in an occluded ear, the VABC indirect path dominates the direct one. At these frequencies also, the objective VBC, VABC and voice OE as well as the subjective VABC OE are negative while the subjective VBC OE is close to zero.

1.1 Introduction

The "hollow voice" occlusion effect (OE) (Killion, 1988) is a known modification of one's own voice perception when wearing a hearing aid or a hearing protector. This effect creates a discomfort that sometimes brings people to remove their device. Hearing aid wearers are then isolated in their silence while hearing protector wearers are at risk of hearing damage. In an attempt towards resolving this problem, the improvement of the characterization of the voice OE is undertaken in the present research.

Different kinds of OE can be found in the literature. Usually the bone conduction (BC) OE is considered (Békésy, 1960a; Lundh, 1986; Wimmer, 1986; Stenfelt et al., 2003; Stenfelt and Reinfeldt, 2007; Reinfeldt et al., 2007). Some studies considered also the voice OE (Lundh, 1986; Wimmer, 1986; Hansen, 1997, 1998; Dillon, 2000). Moreover the OE can be quantified in an objective or subjective manner. These different OEs have been defined in the literature (Tonndorf et al., 1966; Tonndorf, 1972; Hansen, 1997, 1998; Stenfelt et al., 2002; Reinfeldt et al., 2007; Stenfelt and Reinfeldt, 2007; Berger, 1986) and are presented here under a unified terminology that will be used throughout this article.

- **Objective BC OE:** The objective BC OE can be defined as the difference of the sound pressure level (SPL) in the ear canal between occluded and open ear conditions when the subject is submitted to a bone vibrator.
- **Subjective BC OE:** The subjective BC OE can be defined as the difference of the hearing perception between occluded and open ear conditions when the subject is submitted to a bone vibrator.
- **Objective voice OE:** The objective voice OE can be defined as the difference of the SPL in the ear canal between occluded and open ear conditions when the subject is speaking. The subject is supposed to speak at the same level whether the ear is occluded or not.
- **Subjective voice OE:** The subjective voice OE can be defined as the difference of the voice perception between occluded and open ear when the subject is speaking.

Although all these OEs have been defined, not all of them have been quantified. The objective BC OE (Stenfelt et al., 2003; Stenfelt and Reinfeldt, 2007; Reinfeldt et al., 2007; Lundh, 1986; Wimmer, 1986) and the subjective one (Békésy, 1960a; Reinfeldt et al., 2007; Stenfelt and Reinfeldt, 2007) have been quantified by several researchers. The objective voice OE (Dillon, 2000; Hansen, 1997, 1998; Lundh, 1986; Wimmer, 1986) has also been quantified. The subjective voice OE has been examined by Hansen (1997) who had evaluated the annoyance experienced by subjects when they wear their hearing aids. However no real quantification (threshold or loudness) of this effect was done in her study.

The purpose of the present paper is to quantify the subjective voice OE and to analyze its different aspects. The subjective voice OE can hardly be obtained by the classical method of hearing thresholding, but it could be measured by the loudness balance method. As this second possibility is more time consuming and more demanding for the subjects and probably less accurate, a new hearing thresholding measurement has been developed. The voice in the subject's mouth is simulated by means of an external speaker.

The article is organized as follows. In section 1.2 the main physical explanations of the OE are presented for later use in the data analysis. Section 1.3 describes the different internal sound path components involved in the perception of one's own voice and introduces the possible measurements and their associated OEs. Some measurements were previously available in the literature whereas some were designed and implemented in this present research. The new measurement method is presented in section 1.4. The experimental results are in section 1.5. In section 1.6, the voice OE is characterized using the present results and these from the literature. The conclusions and recommendations are drawn in section 3.5.

1.2 Main physical explanations of the OE

The purpose of this section is to briefly present the main physical mechanisms that are generally agreed upon, in the literature, to explain the OE. The upcoming section on the characterization of the voice OE (section 1.6) will refer to these physical mechanisms whenever they can be used to explain the measurement results.

The OE is most generally observed and explained at low frequencies (up to about 2 kHz) (Stenfelt et al., 2003): it is an increase in the perceived and measured sound level in the ear canal. If the external air conduction is not considered, the ear canal SPL is in a large part due to the vibrations of the cartilaginous walls (as proposed by Bárány (1938)) which are located roughly along the external $2/3$ of the ear canal. Occlusion modifies the ear canal acoustic system, preventing at the opening the radiation of sound which is typically more efficient at low frequencies, so more low frequency sounds remain in the ear canal and are transmitted to the inner ear. A higher SPL in the ear canal and a lower auditory threshold can then be observed at low frequencies. This effect has been presented in Tonndorf (1972) and Tonndorf et al. (1966) who have shown that the opened ear canal behaves like a high-pass filter based on a lumped parameter element model. Stenfelt et al. (2003) confirmed Tonndorf's explanation in his 2003 experiments where he showed a positive OE up to 2 kHz. Several authors have also demonstrated that this positive OE is practically eliminated or much lowered when the earplug is deeply inserted, which is attributed to the fact that the cartilaginous part of the ear canal is completely covered by the earplug and only the bony part is radiating at a much lower level than the cartilaginous part (Berger and Kerivan, 1983; Stenfelt and Reinfeldt, 2007).

At high frequencies (above about 2 kHz), the OE is only partially explained. The explanations are based on a comparison of the ear canal acoustic system in open and occluded configurations. In experiments where the ear canal is replaced by an equivalent plastic tube, Stenfelt et al. (2002) demonstrated that the open tube has a quarter wavelength resonance at 2.7 kHz, whereas the closed tube has a half wavelength resonance at 5.5 kHz. Rinne (1855), Politzer (1907-1913) and Huizing (1960) (cf. Tonndorf (1972)) proposed the assumption that the OE could be caused by these resonance modifications of the ear canal. Tonndorf et al. (1966); Tonndorf (1972) and Stenfelt et al. (2003) validated this assumption but only for high frequencies. This has also been obtained by Hansen (1998) with a distributed parameter element model of the ear canal which is represented by a constant section tube. Hence, at high frequencies, the OE could be explained by the facts that suppressing the quarter wavelength resonance at about 2.7 kHz would give a negative OE around this frequency and that a positive OE would appear around the half wavelength resonance at about 5.5 kHz (or above since the ear canal length is reduced by the earplug insertion depth).

1.3 The internal sound path components involved in the perception of one's own voice

Voice production is a complex mechanism involving air moving out of the lungs that induces a fluctuating pressure at the entrance of the vocal tract that will filter this pressure signal to produce an external voice signal (Gelfand, 2004). The aforementioned fluctuating pressure can be produced in two ways (Gelfand, 2004): (i) for voiced sounds that constitute most of the components of speech delivered at a conversation level, the fluctuating pressure at the entrance of the vocal tract is induced mainly through vocal cords vibrations, (ii) for unvoiced sounds that constitute all the components of whispered speech and some components of usual speech such as some consonant sounds, the fluctuating pressure at the entrance of the vocal tract is not induced by vocal cord vibrations but by turbulences generated by flow restrictions. In other words, there are two voice sources: the vocal cords vibrations and the turbulences due to flow restrictions. The vocal cord vibrations source has the particularity of interacting with the air in the larynx and also interacting directly with the bone and associated body structures supporting the vocal cords, so it will be classified in this text as both an airborne source and a structure-borne source. The turbulences that interact only with air will be classified as an airborne source.

From the point of view of voice perception, two main paths can be distinguished from the voice sources to the inner ear (Howell, 1985): the external and the internal paths. These two paths have also been identified by Békésy (1949, 1960a) who found that the amount of acoustic energy transmitted by each of them is of the same order. The external voice perception is only due to the airborne source, whereas the internal voice perception depends on both the structure-borne source and the airborne source. Since the present research aims at characterizing the OE, only the internal path will be studied. In order to easily refer to the various components of this internal path, they have been represented schematically with a bloc diagram in Fig. 1.1.

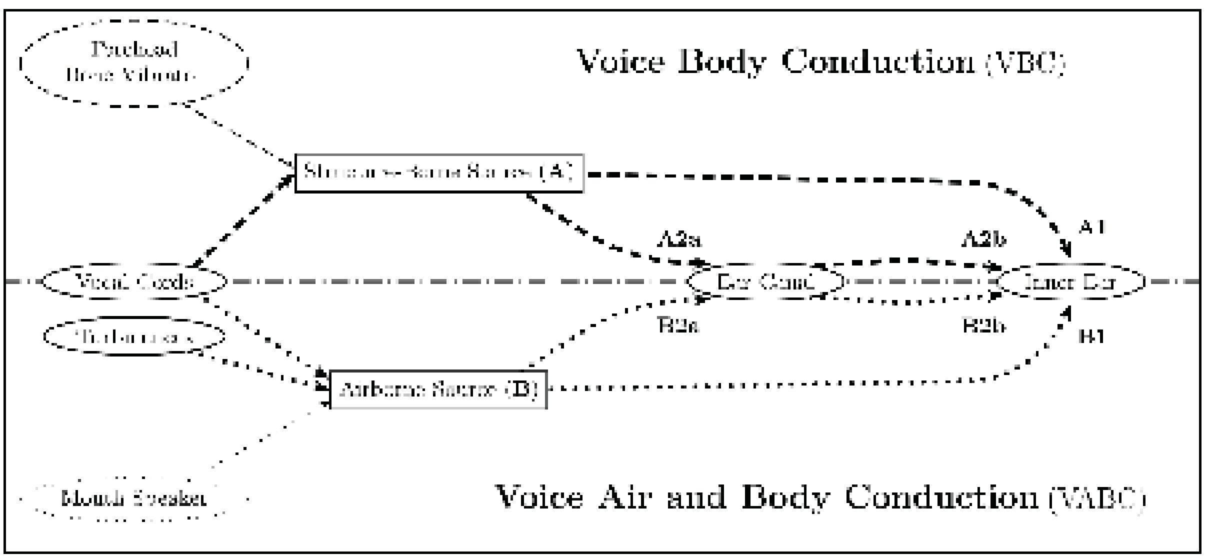

Figure 1.1 Diagram of the internal sound path components involved in the perception of one's own voice.

Following Békésy (1949, 1960a) and Howell (1985) this internal path is separated into two paths according to two possible sources: (i) a structure-borne source (**A**) due to the vocal cord vibrations that excite directly the skull bones, (ii) an airborne source (**B**) (vocal cords interaction with air or turbulences in the larynx) that excites the vocal tract cavities.

The upper half of the bloc diagram contains the first path, the voice body conduction (VBC) path associated with the structure-borne source (**A**). The structure-borne source (**A**) is either the vocal cords (voiced sound production) or a bone vibrator (artificial excitation on the skull, most often the forehead or the mastoid bone) used in many bone excitation experiments reported in the literature (Békésy, 1960a; Lundh, 1986; Wimmer, 1986; Stenfelt et al., 2003; Stenfelt and Reinfeldt, 2007; Reinfeldt et al., 2007). According to the works on BC of Békésy (1960a) and Tonndorf (1972)), this VBC path is subdivided into two paths: (i) a direct path to the inner ear (**A1**), (ii) an indirect path, first to the ear canal (**A2a**) and second, from the ear canal to the inner ear (**A2b**). In naming this path as well as the other one in the lower half of the bloc diagram, the word "body" is used instead of the word "bone" to acknowledge the fact that not only the bones but other human tissues contribute to solid borne sound transmission.

The lower half of the bloc diagram contains the second path, the voice air and body conduction (VABC) path associated with the airborne source (**B**). The airborne source (**B**) is either, for real voice production, the vocal cords interaction with air and the air turbulences in the larynx or, for a path identification experiment developed specifically for the presented research work, an artificial excitation using a mouth speaker (the details on the experimental setup are given in section 1.4). This VABC path can be subdivided, exactly like the VBC path, into a direct path to the inner ear (**B1**) and one indirect path, first to the ear canal (**B2a**) and second, from the ear canal to the inner ear (**B2b**).

For both the VBC and the VABC paths, two kinds of measurements can be performed, a subjective one for the paths reaching the "inner ear" box which represents the subject hearing perception, and an objective one for the paths reaching the "ear canal" box where a microphone can give the SPL value. In the following paragraphs, the four OEs defined in the introduction are expressed in terms of the measurements (for open and occluded ear) of the associated paths.

The subjective voice OE is the perception difference between open and occluded ear associated with both the VBC and the VABC paths when the source is the subject real voice (vocal cords and/or turbulences). Classically, as for the subjective BC OE (Békésy, 1960a; Reinfeldt et al., 2007; Stenfelt and Reinfeldt, 2007), the threshold measurements for open and occluded ear is used to quantify the OE. However, for the subjective voice OE, the standardized threshold measurement protocol (ISO 8253-1, 1989; ISO 8253-2, 1992) cannot be used anymore because the subject will always know when a sound is possibly emitted. A possibility could be to use a loudness balance method. However no account of a loudness balance method for the determination of the subjective voice OE was found in the literature, only qualitative evaluations of this effect have been found in (Hansen, 1997).

The objective voice OE is the ear canal SPL difference between open and occluded ear associated this time with a fraction only of the path associated with the subjective OE presented in the previous paragraph: only the **A2a** path from the VBC and the **B2a** path from the VABC are considered in this case. Some results of the literature (Lundh, 1986; Wimmer, 1986; Hansen, 1997, 1998; Dillon, 2000) will be used in section 1.6 for analysis purpose.

The first two OEs presented (subjective and objective voice OEs) include the two kinds of voice sources: the structure-borne source(**A**) and the airborne source(**B**). In order to investigate separately the OEs associated with each of these two kinds of sources, the real sources can be replaced by an artificial one: a forehead bone vibrator which may be said similar to the structure-borne source part of the vocal cords, or a mouth speaker which may be said similar to the turbulences and the airborne source part of the vocal cords.

In the case of the structure-borne source (**A**), according to Békésy (1949, 1960a), the VBC path can be considered similar to a BC path. Then the VBC OEs can be said equivalent to the BC ones and the literature results can be used. The subjective VBC OE is the hearing threshold difference between open and occluded ear associated with the VBC paths (**A1** in parallel with **A2a** and **A2b**). The objective VBC OE is the ear canal SPL difference between open and occluded ear associated only with the **A2a** path from VBC. Some literature results on these two BC OEs (subjective (Reinfeldt et al., 2007; Berger and Kerivan, 1983; Stenfelt and Reinfeldt, 2007) and objective (Reinfeldt et al., 2007; Stenfelt and Reinfeldt, 2007; Wimmer, 1986; Lundh, 1986)) will be used in section 1.6 for analysis purpose.

In the case of the airborne source (**B**), similarly to the structure-borne source (**A**), we define two OEs: the subjective VABC OE is the hearing threshold difference between open and occluded ear associated with the VABC paths (**B1** in parallel with **B2a** and **B2b**) whereas the objective VABC OE is the ear canal SPL difference between open and occluded ear associated only with the **B2a** path from VABC. As mentioned previously in this section, an experiment is designed using a mouth speaker to simulate the voice airborne source. In this research, a subjective measurement is conducted giving a quantification of the subjective VABC OE. The objective measurement could be considered in future research. The experimental set-up and protocol for this subjective measurement of the VABC OE are presented in the next section.

1.4 Measurement method

The experimental protocol has been examined and accepted by the IRB (Institutional Review Board) of the École de technologie supérieure.

1.4.1 Subjects

Eleven subjects have been chosen for their otologically normal ears and their good hearing using criteria from the ISO 4869-1 standard (ISO 4869-1, 1990): their minimum audible pressure (MAP) must be below 20 dB-HL for frequencies below 2,000 Hz and below 30 dB-HL for frequencies above 3,000 Hz.

For each subject, a pair of Sonomax v3 S2/M2 (small or medium size) custom earplugs from Sonomax Hearing Healthcare Inc. (Montréal Québec) are fitted individually by a certified implementor. The average insertion depth of the earplugs is 13 mm. For easier comparison with data in the literature, this insertion depth can be said equivalent to an approximate average occluded volume value of about 0.54 cm^3, based on an average length of the ear canal of 27 mm and an average diameter of 7 mm (Small and Gales, 1998).

1.4.2 Experimental setup

All tests were realized in an audiometric booth with an Interacoustics AC40 clinical audiometer. The 125, 250, 500, 1000, 2,000, 4,000, and 8,000 Hz frequencies are tested with a warble tone according to the ISO 8253-1,2 standards (ISO 8253-1, 1989; ISO 8253-2, 1992). All tests are realized in open and occluded ears with SonoCustom v3 S2/M2 earplugs.

The experiments are perform in two experimental phases. In the first one, measurements in a diffuse field are realized. In the second one, an acoustic box is used. The flowchart of these experimentals phases is reprensented in Fig. 1.2.

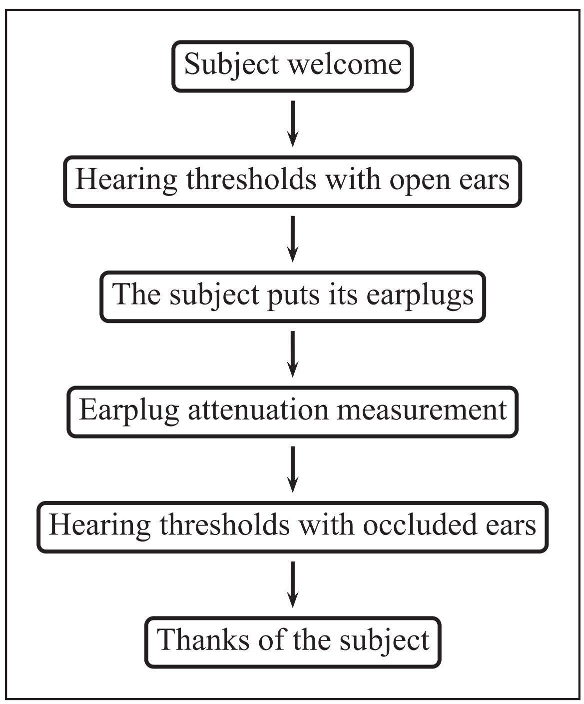

Figure 1.2 Flowchart of each experimental phases

In a first experimental phase, the hearing thresholds with open and occluded ears were measured in a diffuse field in the audiometric booth. This diffuse field was previously calibrated according to the ISO 389-7 standard (ISO 389-7, 1996). The hearing threshold with open and occluded ears measurements give respectively the minimum audible diffuse field in open ears ($\mathrm{MADF_{op}}$) and in occluded ears ($\mathrm{MADF_{oc}}$). The REAT (Real Ear Attenuation at Threshold: $\mathrm{MADF_{op}} - \mathrm{MADF_{oc}}$) is then quantified. This measurement serves to check that the earplug attenuation is within ANSI certification (Industrial Noise Laboratory, 2007).

In a second experimental phase, an acoustic box is used to transmit a sound field directly and only in the mouth of the subject. The acoustic box was designed to provide an acoustic signal in the subject's mouth by making the sound of a loudspeaker box converge to a spirometry filter system with a tubular extremity on which the subject closed his mouth. The experimental set-up including details on the acoustic box elements are presented in Fig. 1.3.

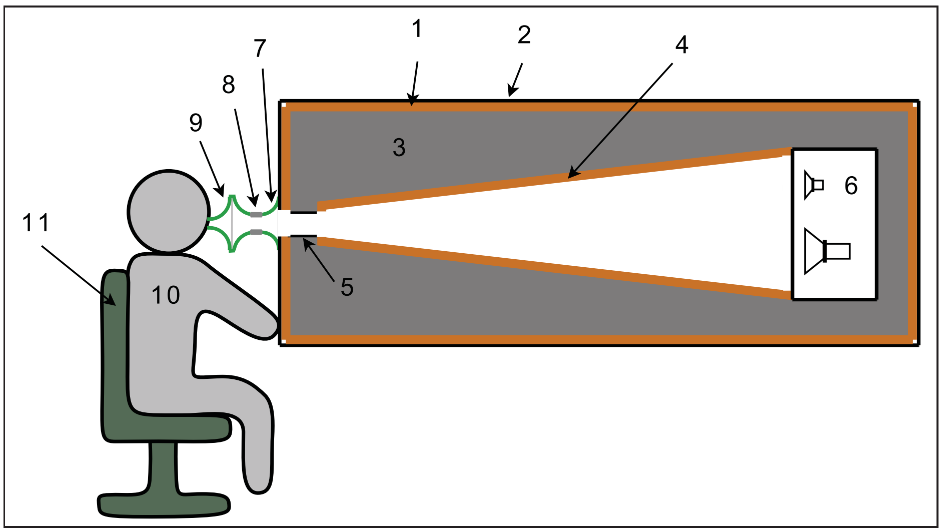

Figure 1.3 Setup of the acoustic box: (1) parallelepipedic box in 1 inch plywood, (2) sound barrier, (3) at least 4 inches of sound absorber, (4) truncated pyramid with rectangular base and square node in 1 inch plywood, (5) PVC pipe of 1,75 inches diameter, (6) speaker, (7) fixed half spirometry filter, (8) spirometry filter adapter, (9) individual spirometry filter (bacterial / viral), (10) subject, (11) adjustable seat. (color online)

Since there is no standard available for the calibration of this new measurement procedure, it was chosen, somewhat arbitrarily, to apply the ISO 389-7 standard (ISO 389-7, 1996) for diffuse field to calibrate the sound field at the output of the ergo-filter closed with a microphone adapter. This choice has no effect on the OE measurement since it is a threshold difference, it only affects the hearing threshold value as represented on the results in Fig. 1.5 of the next section. The hearing thresholds with open and occluded ears are measured with this acoustic box set-up in the audiometric booth. The minimum audible mouth pressure in open ears ($\mathrm{MAMP_{op}}$) and in occluded ears ($\mathrm{MAMP_{oc}}$) are obtained. They serve to determine the OE associated with the VABC path.

Because each experimental phase is fairly demanding for the subjects, it was chosen to perform the two experimental phases at two different moments. Hence the earplug was removed and reinserted and other factors might affect earplug attenuation. Consequently, in order to check that earplug attenuation had not changed significantly between the two phases, a noise reduction measurement of earplugs was carried out before each experiment with occluded ears. This is presented in the next section.

1.4.3 Noise reduction measurement

The noise reduction provided by earplugs is measured in a 85 dB-SPL free field pink noise by means of two dual microphone probes (produced by Sonomax Hearing Healthcare Inc.) which measure the SPL inside (p_{meas}) and outside (p_{ref}) the earplug for both ears. In Fig. 1.4, a Sonomax v3 S2/M2 custom earplug with an inserted dual microphone probe is drawn schematically.

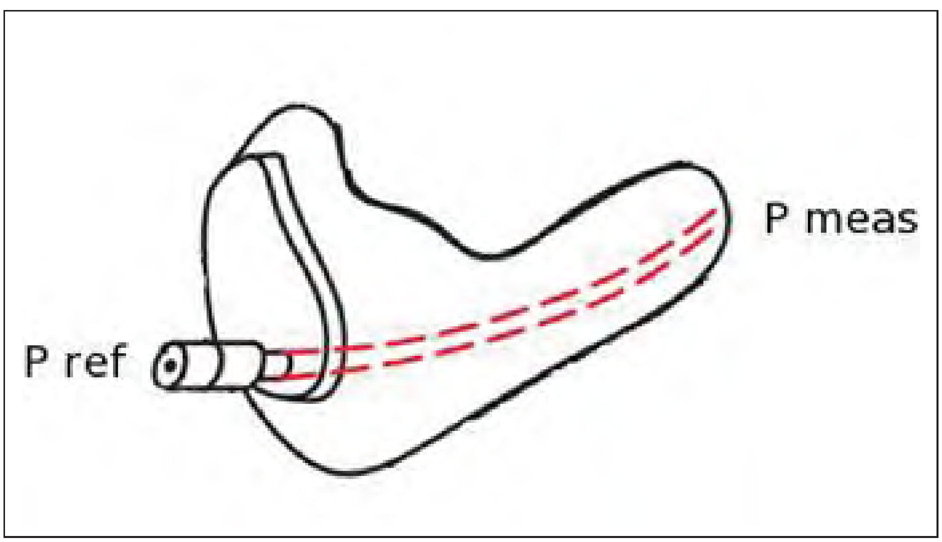

Figure 1.4 Sonomax v3 S2/M2 custom earplug with an inserted dual microphone probe (color online).

The noise reduction measurement protocol which follows is the second scenario described by Voix (2006) and Voix and Laville (2009) with adapted notations. The measured noise reduction $\mathrm{NR_{meas}}$ of earplugs realized when the subject wears his earplugs is defined by equation 1.1.

$$\mathrm{NR_{meas}} = 20\ \log_{10}\left(\frac{p_{\mathrm{ref}}}{p_{\mathrm{meas}}}\right) \tag{1.1}$$

This measure must be corrected by the transfer function $\tilde{\mathrm{NR}}$ of the dual microphone probe inserted in the earplug when they are not worn and submitted to a uniform acoustic pressure field:

$$\tilde{\mathrm{NR}} = 20\ \log_{10}\left(\frac{\tilde{p}_{\mathrm{meas}}}{\tilde{p}_{\mathrm{ref}}}\right) \tag{1.2}$$

The corrected noise reduction $\mathrm{NR_c}$ (cf. equation 1.3) is obtained by subtracting this transfer function from the measured noise reduction

$$\mathrm{NR_c} = \mathrm{NR_{meas}} - \tilde{\mathrm{NR}} \tag{1.3}$$

A binaural corrected noise reduction $\mathrm{NR_{cb}}$ is then obtained by the equation 1.4 with MAP the minimum audible pressure under headphone of the subject.

$$\begin{cases} if\ \Big(\mathrm{NR_c(left)} + \mathrm{MAP(left)}\Big) < \Big(\mathrm{NR_c(right)} + \mathrm{MAP(right)}\Big) \\ \qquad \mathrm{NR_{cb}} = \mathrm{NR_c(left)} \\ if\ \Big(\mathrm{NR_c(left)} + \mathrm{MAP(left)}\Big) > \Big(\mathrm{NR_c(right)} + \mathrm{MAP(right)}\Big) \\ \qquad \mathrm{NR_{cb}} = \mathrm{NR_c(right)} \\ if\ \Big(\mathrm{NR_c(left)} + \mathrm{MAP(left)}\Big) = \Big(\mathrm{NR_c(right)} + \mathrm{MAP(right)}\Big) \\ \qquad \mathrm{NR_{cb}} = \min\Big(\mathrm{NR_c(left)},\ \mathrm{NR_c(right)}\Big) \end{cases} \tag{1.4}$$

1.5 Experimental results

Figure 1.5 depicts the main average experimental results: the minimum audible diffuse field in open ears ($\mathrm{MADF_{op}}$) and in occluded ears ($\mathrm{MADF_{op}}$), and the minimum audible mouth pressure in open ears ($\mathrm{MAMP_{op}}$) and in occluded ears ($\mathrm{MAMP_{oc}}$).

In order to validate our experimental results, the noise attenuation of earplugs is examined in a first part. In a second part, the objective VABC OE is presented.

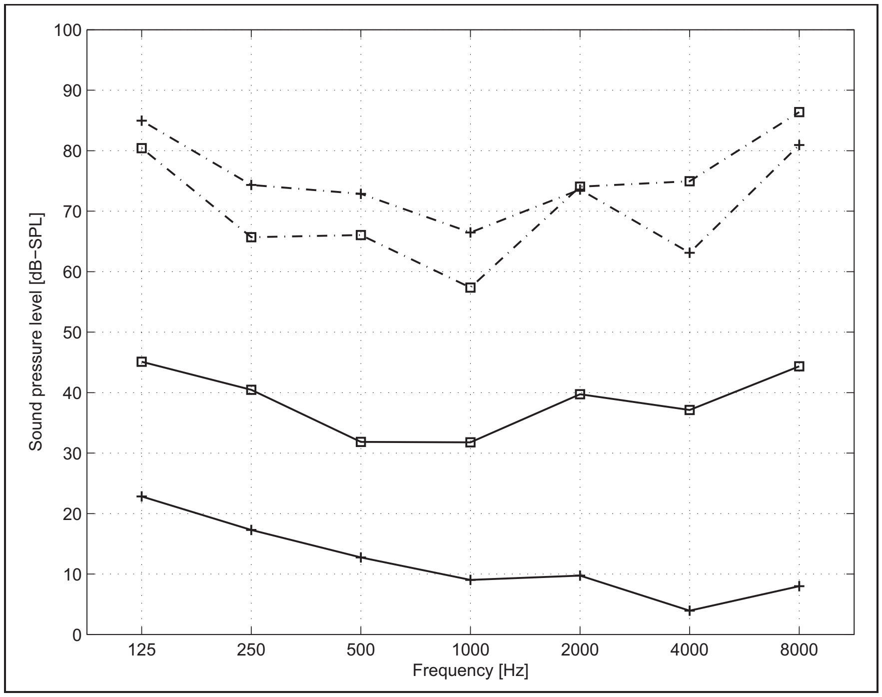

Figure 1.5 Average $\mathrm{MADF_{op}}$ ($-+$), average $\mathrm{MADF_{oc}}$ ($-\Box$), average $\mathrm{MAMP_{op}}$ ($-\cdot+$), and average $\mathrm{MAMP_{oc}}$ ($-\cdot\Box$).

1.5.1 Noise attenuation of earplugs

In Fig. 1.6 are plotted the experimental average and standard deviation REAT (real ear attenuation threshold: $\mathrm{MADF_{op}} - \mathrm{MADF_{oc}}$) obtained in diffuse field as well as the certified ANSI REAT (Industrial Noise Laboratory, 2007) (average $\pm$ two standard deviation and standard deviation) for the Sonomax v3 S2/M2 custom earplugs. As can be seen in Fig. 1.6 the experimental REAT results are in agreement with the ANSI certification ones.

As mentioned previously, the NR values were obtained before each experimental phase with occluded ears to check that the earplugs' attenuation had not changed significantly between the two phases. The delta between $\mathrm{NR_{cb}}$ obtained in the two experimental phases is presented in Fig. 1.7. At low frequencies there are practically no differences. At high frequencies the delta is slightly higher, but still acceptable, especially because the fit variability of earplugs is high at high frequencies. The small differences obtained confirmed that the subjects have placed their earplugs in the same manner during the two experimental phases. As the earplugs have

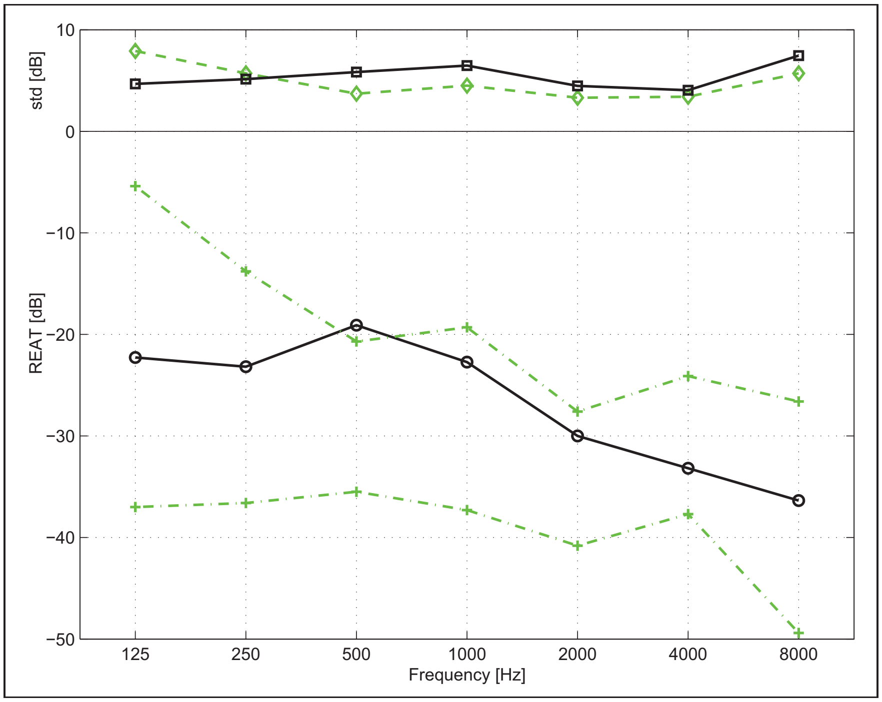

Figure 1.6 REAT experimental results [average (–◯), std (–□)] and REAT ANSI results (Industrial Noise Laboratory, 2007) [average ± 2 std (– – +), std (– – ◊)] (color online).

been well fitted during the diffuse field experiments, they also have been well fitted during the experiments with the acoustic box.

1.5.2 Subjective VABC OE

In Fig. 1.8 are plotted the average and the standard deviation of the subjective VABC OE which can be quantified by the difference between $\mathrm{MAMP}_{\mathrm{op}}$ and $\mathrm{MAMP}_{\mathrm{oc}}$. The subjective VABC OE is positive at low frequencies (below 2,000 Hz) and negative at high frequencies.

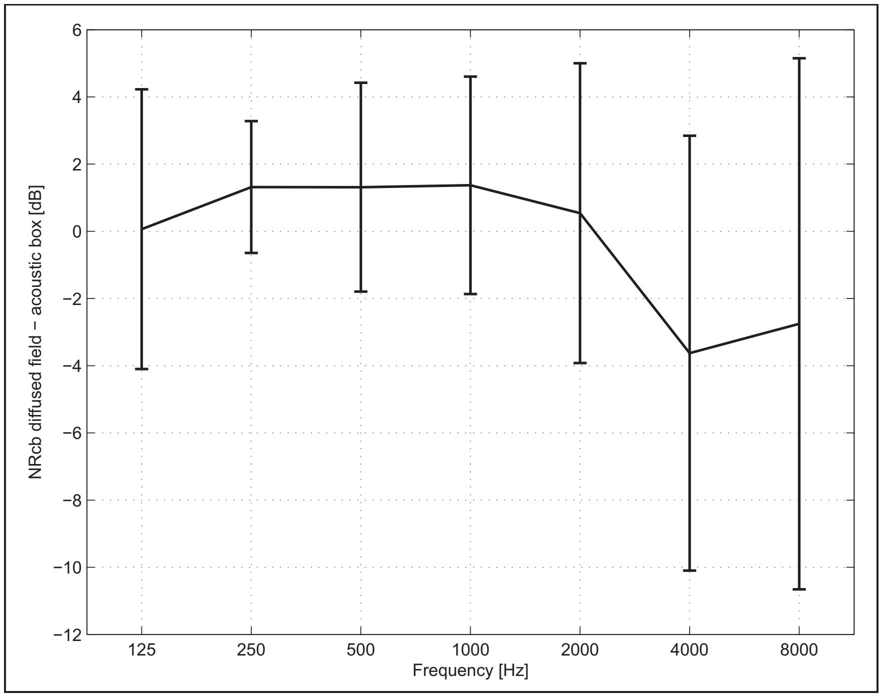

Figure 1.7 Delta between $\mathrm{NR_{cb}}$ in diffused field and acoustic box experiments.

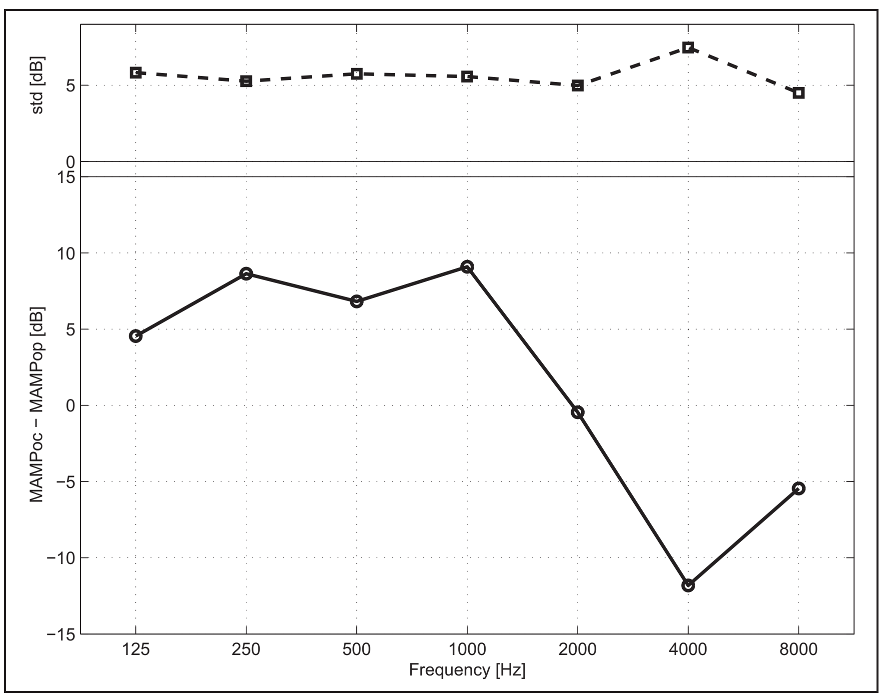

Figure 1.8 Subjective VABC OE ($\mathrm{MAMP_{oc}} - \mathrm{MAMP_{op}}$) [average ($-\bigcirc$), std ($- - \square$)]

1.6 Characterization of the voice OE using experimental results from the presented research work and from the literature

For each of the OEs considered here (the subjective and objective BC OE, the objective voice OE from the literature and the subjective VABC OE from the presented experimental work), the analysis will be separated in two frequency domains, each one in a separate subsection: low frequencies (below 1500-2000 Hz) and high frequencies (above 1500-2000 Hz). For each frequency domain, firstly we determine which path dominates in occluded and opened ears. Secondly, we determine how occlusion generates path modifications leading to a positive, negative or null OE. Thirdly, we verify that the obtained results are in agreement with the main physical mechanism, described in section 1.2, which can only be applied to the indirect paths (**A2a** and **B2a**). Information on path identification and OE will only be given when a conclusion can be reached.

In order to easily determine the relative weight of each path in every case, for the rest of this article, the name of the path, for example **B1**, will also be used to denote the amount of acoustic energy that the path brings to the end point, so the path **B1** brings the energy **B1** to the inner ear which receives the sum of the energies **B1**, **B2b**, **A1**, and **A2b**. At the inner ear, energy should be understood as perceived energy by the subject, whereas at the ear canal, the energy is the real acoustic energy which can be measured as SPL with a microphone.

Firstly, we analyze the subjective VABC OE we quantified in our experiments. Then we consider successively the OEs which have been already quantified in the literature (subjective and objective BC OEs and objective voice OE). Finally we synthesize the results obtained for all these OEs.

1.6.1 Subjective VABC OE

According to Fig. 1.1, the VABC path corresponds to the paths **B1** and **B2** (**B2a** then **B2b**). The analyses that follow aim at determining the relative contribution of these components in both open and occluded ears according to the results presented in Fig. 1.8.

1.6.1.1 At low frequencies

At low frequencies, the positive OE of the order of $+5$ to $+10$ dB means that the energy transmitted by the sum of the **B1** and **B2** paths is much more important in occluded ear than in opened ear: $\mathbf{B1}_{\mathrm{oc}} + \mathbf{B2}_{\mathrm{oc}} \gg \mathbf{B1}_{\mathrm{op}} + \mathbf{B2}_{\mathrm{op}}$. As mentioned in section 1.2, the occlusion has an effect on the ear canal vibratory and acoustic system, hence, a reasonable assumption is that occlusion will affect only the **B2** path and leave unchanged the **B1** path, so that $\mathbf{B1}_{\mathrm{oc}} \approx \mathbf{B1}_{\mathrm{op}}$. This equation put into the first inequality leads to $\mathbf{B2}_{\mathrm{oc}} \gg \mathbf{B2}_{\mathrm{op}}$ and, furthermore, on the left hand-side of the inequality $\mathbf{B2}_{\mathrm{oc}} \gg \mathbf{B1}_{\mathrm{oc}}$. Indeed, if we had $\mathbf{B2}_{\mathrm{oc}} \ll \mathbf{B1}_{\mathrm{oc}}$, then we would have $\mathbf{B1}_{\mathrm{oc}} \gg \mathbf{B1}_{\mathrm{oc}} + \mathbf{B2}_{\mathrm{op}}$ which is false. So the assumption $\mathbf{B2}_{\mathrm{oc}} \ll \mathbf{B1}_{\mathrm{oc}}$ is false. Otherwise, if we had $\mathbf{B2}_{\mathrm{oc}} \approx \mathbf{B1}_{\mathrm{oc}}$, then we would have $\mathbf{B1}_{\mathrm{oc}} \cdot 2 \gg \mathbf{B1}_{\mathrm{oc}} + \mathbf{B2}_{\mathrm{op}}$ which is false. So the assumption $\mathbf{B2}_{\mathrm{oc}} \approx \mathbf{B1}_{\mathrm{oc}}$ is false.

In other words, for the VABC perception at low frequencies we can conclude:

- Path identification: in an occluded ear the indirect path **B2** transmits more energy than the direct path **B1**;
- OE: a positive OE results from the fact that occlusion increases the energy transmitted by the **B2** path which dominates the **B1** path.

As the indirect path **B2** dominates, this is in agreement with the low frequency physical mechanisms presented in section 1.2. In open ear, the low frequencies ear canal wall radiations are propagated out of the ear. In occluded ear, these low frequencies are blocked inside and so the SPL at low frequencies is increased.

1.6.1.2 At high frequencies

At high frequencies, the negative OE of the order of -5 to -12 dB means that the energy transmitted by the sum of the **B1** and **B2** paths is much more important in opened ear than in occluded ear: $\mathbf{B1}_{\mathrm{oc}} + \mathbf{B2}_{\mathrm{oc}} \ll \mathbf{B1}_{\mathrm{op}} + \mathbf{B2}_{\mathrm{op}}$. As in the case of the low frequencies the same assumption can be made to the effect that occlusion will affect only the **B2** path and leave unchanged the **B1** path, so that $\mathbf{B1}_{\mathrm{oc}} \approx \mathbf{B1}_{\mathrm{op}}$. This equation put into the first inequality leads to $\mathbf{B2}_{\mathrm{oc}} \ll \mathbf{B2}_{\mathrm{op}}$ and, furthermore, on the right hand-side of the inequality $\mathbf{B2}_{\mathrm{op}} \gg \mathbf{B1}_{\mathrm{op}}$.

In other words, for the VABC perception at high frequencies we can conclude:

- Path identification: in an open ear the indirect path **B2** transmits more energy than the direct path **B1**;
- OE: a negative OE results from the fact that occlusion decreases the energy transmitted by the **B2** path which dominates the **B1** path.

As the indirect path **B2**, this is in agreement with the high frequency physical mechanism described in section 1.2. The canal occlusion causes modifications of its resonances. In opened ear, the first ear canal resonance is about 2.7 kHz. In occluded ear, this resonance disappears and is replaced by a higher one (5.5 kHz or above).

1.6.2 Subjective BC OE

According to Fig. 1.1, the subjective BC corresponds to the paths **A1** and **A2** (**A2a** then **A2b**). In Fig. 1.9 seven subjective BC OEs from the literature are presented and our subjective VABC OE is also represented as it will be represented on all the following graphs for reference. Three (one from Reinfeldt et al. (2007) and two from Stenfelt and Reinfeldt (2007)) of the seven presented measurements cover the whole frequency range from 125 to 8000 Hz, they have been chosen because they correspond to insertion depths (18 mm, 15 mm and 7 mm) close to the one in this study (13 mm) and, for these three measurements, the objective BC OE has also been quantified in the same conditions (and will be presented in the next section). The other four measurements cover a more limited frequency (125 to 2000 Hz), but they have been chosen because they are four measurements from the same author (Berger and Kerivan, 1983) that cover a wide range of insertion depth (indicated in terms of occluded ear canal volumes of respectively 0.2, 0.5, 0.6, and 0.8 cm^3).

1.6.2.1 At low frequencies

As can be seen in Fig. 1.9 at low frequencies (below 2000 Hz), the subjective BC OE level is much dependent on the earplug insertion and on the kind of earplug. However, the subjective BC OE is globally positive just like in the case of the subjective VABC OE, so the same relations that have been introduced previously for the VABC case will be valid here, after the letter

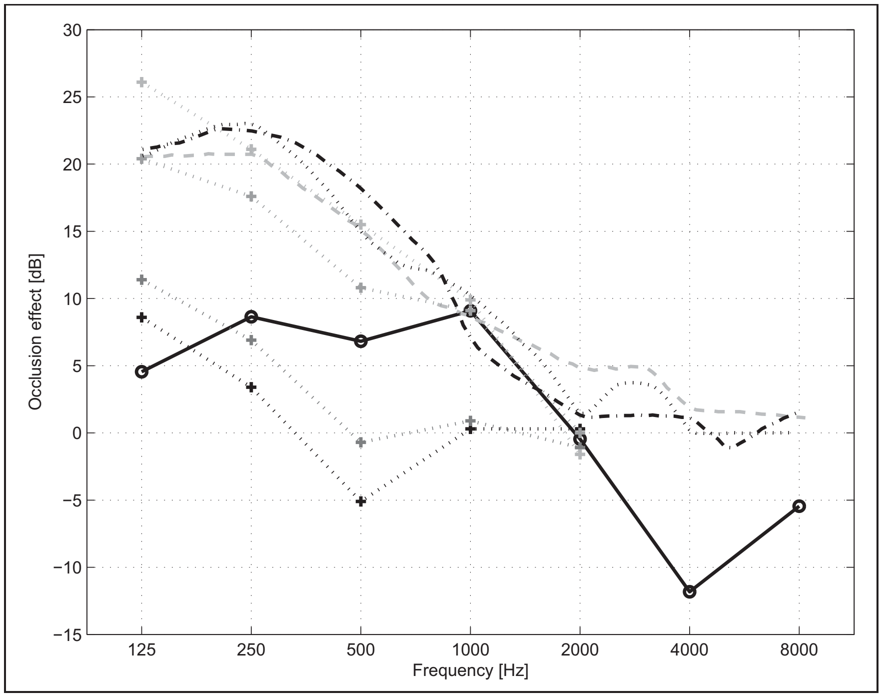

Figure 1.9 Subjective BC OEs of the literature [this study (−◯ black); Reinfeldt et al. (2007): (−− gray) occlusion with 18 mm insertion of E-A-R Classics and bone excitation at the forehead; Berger and Kerivan (1983) bone excitation at the forehead: ($\cdots +$ black) E-A-R plug with 0.2 cm^3 occluded volume, ($\cdots +$ dark gray) E-A-R plug with 0.5 cm^3 occluded volume, ($\cdots +$ medium gray) V-51R plug with 0.6 cm^3 occluded volume, ($\cdots +$ light gray) E-A-R plug with 0.8 cm^3 occluded volume; Stenfelt and Reinfeldt (2007) bone excitation at the forehead: ($\cdots$ black) occlusion with 7 mm insertion of foam earplug, (−· black) occlusion with 15 mm insertion of foam earplug].

B is replaced by the letter A: $\mathbf{A1}_{\mathrm{oc}} + \mathbf{A2}_{\mathrm{oc}} \gg \mathbf{A1}_{\mathrm{op}} + \mathbf{A2}_{\mathrm{op}}$ and, with the reasonable assumption that occlusion will affect only the **A2** path and leave unchanged the **A1** path, $\mathbf{A1}_{\mathrm{oc}} \approx \mathbf{A1}_{\mathrm{op}}$, this will yield $\mathbf{A2}_{\mathrm{oc}} \gg \mathbf{A2}_{\mathrm{op}}$ and $\mathbf{A2}_{\mathrm{oc}} \gg \mathbf{A1}_{\mathrm{oc}}$.

In other words, for the VBC perception at low frequencies we can conclude:

- Path identification: in an occluded ear the indirect path **A2** transmits more energy than the direct path **A1**;
- OE: a positive OE results from the fact that occlusion increases the energy transmitted by the **A2** path which dominates the **A1** path.

For the same reasons as in the case of the subjective VABC OE, as the indirect path (**A2** instead of **B2**) dominates, this is in agreement with the low frequency physical mechanisms presented in section 1.2.

1.6.2.2 At high frequencies

At high frequencies, the fact that the OE is close to zero means that the simultaneous conduction through the **A1** and **A2** paths in the open and occluded ear cases are approximatively equivalent: $\mathbf{A1}_{\mathrm{oc}} + \mathbf{A2}_{\mathrm{oc}} \approx \mathbf{A1}_{\mathrm{op}} + \mathbf{A2}_{\mathrm{op}}$. With the same assumption $\mathbf{A1}_{\mathrm{oc}} \approx \mathbf{A1}_{\mathrm{op}}$ as in the low frequency case, there are two possibilities: the first one is $\mathbf{A2}_{\mathrm{oc}} \approx \mathbf{A2}_{\mathrm{op}}$, the second one is $\mathbf{A1}_{\mathrm{oc}} \gg \mathbf{A2}_{\mathrm{oc}}$ and $\mathbf{A1}_{\mathrm{op}} \gg \mathbf{A2}_{\mathrm{op}}$. The choice will be made in section 1.6.4 using the results from the objective BC OE presented in the next section.

1.6.3 Objective BC OE

According to figure 1.1, the objective BC corresponds to the path **A2a**. In Fig. 1.10 four objective BC OEs from the literature are presented with our subjective VABC OE. As mentioned in section 1.6.2, three of the objective BC OE measurements presented here (one from Reinfeldt et al. (2007) and two from Stenfelt and Reinfeldt (2007)) are issued from the same studies and use the same earplugs as three of the subjective BC OE measurements presented in Fig. 1.9. The fourth one (Lundh (1986), Wimmer (1986)) is included because in their study an objective voice OE has also been quantified in the same conditions (it will be presented in section 1.6.5).

1.6.3.1 At low frequencies

At low frequencies, the OE is positive, so $\mathbf{A2a}_{\mathrm{oc}} \gg \mathbf{A2a}_{\mathrm{op}}$. In other words, for the objective VBC at low frequencies we can conclude:

- OE: a positive OE results from the fact that occlusion increases the energy transmitted by the **A2a** path.

Like for the subjective BC OE, the objective one is in agreement with the low frequency physical mechanism described in section 1.2.

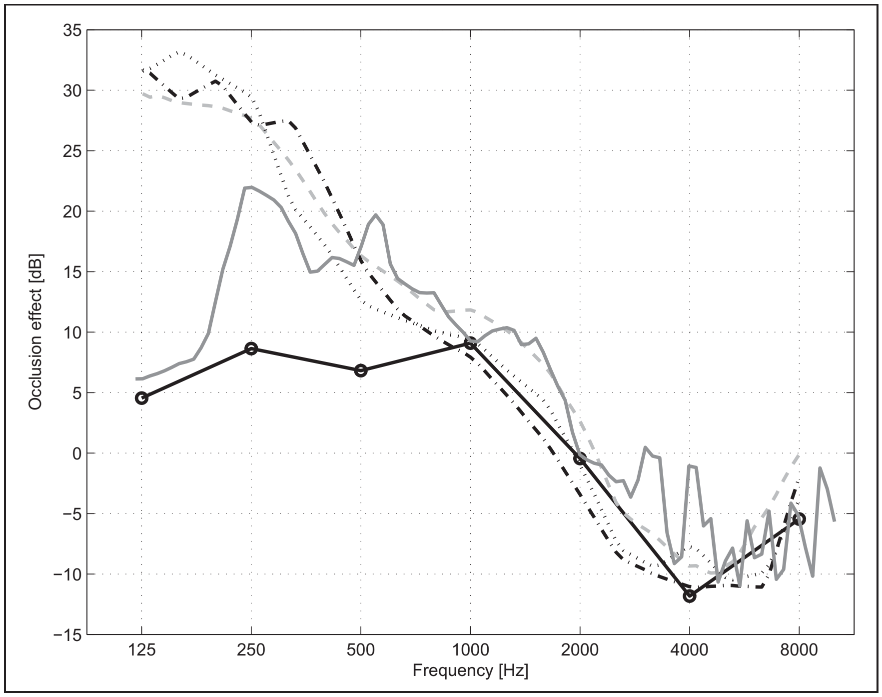

Figure 1.10 Objective BC OEs of the literature [this study (−◯ black); Reinfeldt et al. (2007): (−− gray) occlusion with 18 mm insertion of E-A-R Classics and bone excitation at the forehead; Stenfelt and Reinfeldt (2007) bone excitation at the forehead: (· · · black) occlusion with 7 mm insertion of foam earplug, (−· black) occlusion with 15 mm insertion of foam earplug; Lundh (1986), Wimmer (1986): (− dark gray) occlusion with full earmould impression and bone excitation at the contra-lateral mastoid].

1.6.3.2 At high frequencies

At high frequencies, the OE is negative, so $\mathbf{A2a}_{\mathrm{oc}} \ll \mathbf{A2a}_{\mathrm{op}}$. In other words, for the objective VBC at high frequencies we can conclude:

- OE: a negative OE results from the fact that occlusion decreases the energy transmitted by the **A2a** path.

For the same reasons as for the subjective VABC OE, these results are in agreement with the high frequency physical mechanism described in section 1.2.

1.6.4 BC OE: Integration of physiological noise masking effect and synthesis

In this section, the relations derived in the previous sections on the subjective and objective BC OEs are brought together to characterize as completely as possible the path involved in the BC OE. An analysis of the physiological noise masking effect (PNME) that occurs when the subject is wearing a hearing protector is presented in a first sub-section. This analysis will be used to choose between the two possibilities which have been found for the subjective BC OE at high frequencies (cf. section 1.6.2.2). The two last sub-sections present the general results for BC OE.

1.6.4.1 Physiological noise masking effect

The physiological noise (PN) includes all noises for which the source is in the body (for example breath, heartbeats). The PNME is the masking due to this PN. This effect is more important in occluded ear than in open ear. When the subject is wearing a hearing protector, the low frequency increases in SPL due to the ear canal occlusion affects the PN. The maximum level of the PN under diverse hearing protectors is less than 35 dB SPL around 80 Hz according to Berger and Kerivan (1983), then it decreases with frequencies to reach 20 dB SPL at 250 Hz. The PN in opened ear is lower; in his experiments, Berger and Kerivan (1983) found that it was even lower than the instrumentation noise of his measurement system (about 24 dB SPL at 80 Hz). For both open and occluded ear, the PNME can be written as $\mathbf{A2} = \mathbf{A2b} = \mathbf{A2a} + \mathbf{A2PNME}$ where **A2PNME** represents the PNME that affects the **A2** pathway. The masking effect due to a 250 Hz pure tone at 40 dB SPL is present below 1 kHz and its effect is equivalent to the one due to a narrow band noise (Gelfand (2004) and Ehmer (1959)). Consequently, the **A2PNME** decreases the importance of the **A2** at low frequencies (a little in opened ear $\mathbf{A2PNME}_{op} \leq 0$ and more in occluded ear $\mathbf{A2PNME}_{oc} < 0$), whereas it has nearly no effect at high frequencies ($\mathbf{A2PNME} \approx 0$). We can therefore write $\mathbf{A2} < \mathbf{A2a}$ at low frequencies and $\mathbf{A2} \approx \mathbf{A2a}$ at high frequencies.

1.6.4.2 At low frequencies

At low frequencies, from the analyses of both the subjective and the objective BC OEs as well as the physiological masking effect analysis, the following conclusions can de derived:

- Path identification: From the subjective BC OE, it was found that in an occluded ear the indirect path **A2** transmits more energy than the direct path **A1** ($\mathbf{A2}_{\mathrm{oc}} \gg \mathbf{A1}_{\mathrm{oc}}$);
- OE: A positive OE results for both subjective and objective BC OE because occlusion increases the energy transmitted by the **A2a** path ($\mathbf{A2a}_{\mathrm{oc}} \gg \mathbf{A2a}_{\mathrm{op}}$) as well as the energy transmitted by the **A2** path ($\mathbf{A2}_{\mathrm{oc}} \gg \mathbf{A2}_{\mathrm{op}}$) which dominates the **A1** path . The PNME analysis showed that the perceived energy (hearing threshold) will be higher than the ear canal measured energy (SPL), hence $\mathbf{A2}_{\mathrm{op}} \leq \mathbf{A2a}_{\mathrm{op}}$, $\mathbf{A2}_{\mathrm{oc}} < \mathbf{A2a}_{\mathrm{oc}}$.

As indicated in sections 1.6.2.1 and 1.6.3.1 for the subjective and objective BC OE, these conclusions are in agreement with the low frequency physical mechanism presented in section 1.2.

1.6.4.3 At high frequencies

At high frequencies, the relation $\mathbf{A2a}_{\mathrm{oc}} < \mathbf{A2a}_{\mathrm{op}}$ from the objective BC OE and the relation $\mathbf{A2} \approx \mathbf{A2a}$ from the PNME give: $\mathbf{A2}_{\mathrm{oc}} < \mathbf{A2}_{\mathrm{op}}$. This invalidates the first of the two possibilities found in section 1.6.2.2 that requires $\mathbf{A2}_{\mathrm{oc}} \approx \mathbf{A2}_{\mathrm{op}}$, hence validating the second solution: $\mathbf{A1}_{\mathrm{oc}} \gg \mathbf{A2}_{\mathrm{oc}}$ and $\mathbf{A1}_{\mathrm{op}} \gg \mathbf{A2}_{\mathrm{op}}$. The conclusions for the BC OE at high frequencies are:

- Path identification: in open ear and in occluded ear the direct path **A1** is dominant.
- OE: There is no significant subjective OE because the direct path **A1** is dominant over the indirect path **A2** in both open and occluded ears. However there is a negative objective OE because the indirect path **A2a** transmits less energy when the ear is occluded than when the ear is open. Moreover the indirect path **A2** is more important objectively and subjectively in opened ear than in occluded ear.

As explained for the objective BC OE in section 1.6.3.2, these results are in agreement with the high frequency physical mechanism presented in section 1.2.

1.6.5 Objective voice OE

According to Fig. 1.1, the objective BC corresponds to the sum of the paths **A2a** and **B2a**. In Fig. 1.11 three objective voice OEs from the literature are presented with our subjective VABC OE. As mentioned in section 1.6.3, the objective voice OE measurement from Lundh (1986) and Wimmer (1986) is issued from the same article as the objective BC OE shown in Fig. 1.10. Two others measurements have been found in the literature and are presented here.

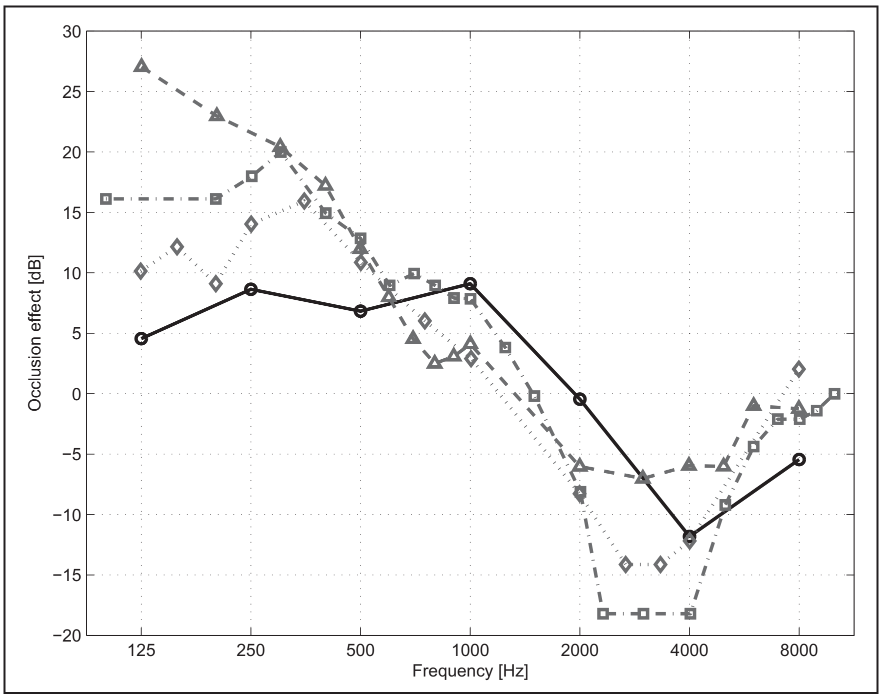

Figure 1.11 Objective voice OEs of the literature [this study (−◯ black); Lundh (1986), Wimmer (1986): (− − △ dark gray) occlusion with earmould impression and bone excitation at the contra-lateral mastoid; Thorup (1996) according to Hansen (1997, 1998): (− · □ dark gray) occlusion with full concha earmould; May (1992) according to Dillon (2000) and Hansen (1997, 1998): (· · · ◊ dark gray) occlusion with unvented skeleton earmould].

1.6.5.1 At low frequencies

At low frequencies, the OE is positive, so $\mathbf{A2a}_{\mathrm{oc}} + \mathbf{B2a}_{\mathrm{oc}} > \mathbf{A2a}_{\mathrm{op}} + \mathbf{B2a}_{\mathrm{op}}$. In other words, for the objective voice at low frequencies we can conclude:

- OE: a positive OE results from the fact that occlusion increases the energy transmitted by the **A2a** and **B2a** paths.

As for the aforementioned low frequencies OEs, this conclusion is in agreement with the low frequency physical explanations presented in section 1.2.

1.6.5.2 At high frequencies

At high frequencies, the OE is negative, so $\mathbf{A2a}_{\mathrm{oc}} + \mathbf{B2a}_{\mathrm{oc}} < \mathbf{A2a}_{\mathrm{op}} + \mathbf{B2a}_{\mathrm{op}}$. In other words, for the objective voice at high frequencies we can conclude:

- OE: a negative OE results from the fact that occlusion decreases the energy transmitted by the **A2a** and **B2a** paths.

Like for previously presented high frequencies OEs, this result is in agreement with the high frequency physical mechanisms described in section 1.2.

1.6.6 Voice OE: Synthesis

In this section the previous specific OEs (from the literature and from the present research) are analyzed simultaneously using the diagram presented in Fig. 1.1. The VBC, the VABC and the voice conduction path (which includes the VBC and VABC paths) and their associated OEs are addressed successively in this section and the conclusions obtained are also presented in Table 1.1 where a question mark is placed in elements for which no answer has been found.

Table 1.1 Main results about path identification and OE for the VBC, the VABC and the voice conduction path

			VBC	VABC	Voice
Low frequencies	Path	Op. ear	?	?	?
		Oc. ear	Indirect $\gg$ Direct	Indirect $\gg$ Direct	Indirect $\gg$ Direct
	OE	Obj.	$\mathrm{OE} > 0$	$\mathrm{OE} > 0$	$\mathrm{OE} > 0$
		Subj.	$\mathrm{OE} > 0$	$\mathrm{OE} > 0$	$\mathrm{OE} > 0$
High frequencies	Path	Op. ear	Direct $\gg$ Indirect	Indirect $\gg$ Direct	?
		Oc. ear	Direct $\gg$ Indirect	?	?
	OE	Obj.	$\mathrm{OE} < 0$	$\mathrm{OE} < 0$	$\mathrm{OE} < 0$
		Subj.	$\mathrm{OE} \approx 0$	$\mathrm{OE} < 0$	?

1.6.6.1 At low frequencies

At low frequencies, all the measured OEs previously presented are found to be positive.

For the VBC path, the subjective measurement lead to the conclusion that in the case of the occluded ear, the indirect path **A2** through the ear canal transmits more energy than the direct path **A1** to the inner ear. On the other hand, the positive subjective and objective OEs result from the energy increase of the indirect path **A2** (and from the previously mentioned fact that **A2** dominates **A1** for subjective measure) in the occluded ear case. Furthermore, the subjective OE is lower than the objective one due to PNME.

For the VABC path, as for the VBC one, the subjective measurement shows that the indirect path **B2** through the ear canal transmits more energy to the inner ear than the direct path **B1** in the occluded ear case. On the other hand, the positive subjective OE effect results from the energy increase of the indirect path **B2** and from the aforementioned fact that **B2** dominates **B1** in the occluded ear case. This subjective VABC OE is fairly constant with frequency and lower than the subjective VBC OE which is inversely proportional to frequency. This difference cannot be readily explained. Two assumptions are proposed. The first one is the different natures of the VABC path (which starts with an air-solid coupling) and the VBC path (which starts with a solid-solid coupling). The second one is the different kinds of earplug used in the considered studies. As far as the objective VABC OE is concerned, although it was not measured, it should be also positive. In fact as the subjective VBC OE is lower than the objective one, the subjective VABC OE should be lower than the objective one.

For the voice conduction path, the objective voice OE is positive, which is justified by the fact that it includes the objective VBC OE and the objective VABC OE which are both positive. For an equivalent reason, even though no subjective voice OE measurements have been done, we can deduce that the subjective voice OE should be positive. As far as the path identification, we previously mention that the indirect path dominates the direct one for the VBC (**A2** $\gg$ **A1**) and VABC (**B2** $\gg$ **B1**) paths. As the voice conduction path is composed of both the VBC and

VABC paths, we can deduce that the indirect path (**A2** + **B2**) should also dominate the direct one (**A1** + **B1**) for the voice conduction path.

1.6.6.2 At high frequencies

At high frequencies, the measured OEs previously presented are found to be either approximatively null or negative.

For the VBC path, the subjective and objective measurement lead to the conclusion that in both cases of the occluded ear and the open ear, the direct path **A1** to the inner ear transmits more energy than the indirect path **A2** through the ear canal. This path identification results in practically no subjective VBC OE as the direct path **A1** to the inner ear is nearly not affected by occlusion. However there is a negative objective VBC OE because the SPL measurement in the ear canal involves only the indirect path **A2a** whose energy appeared to be lowered by occlusion.

For the VABC path, the subjective measurement conducted in this study shows a negative OE instead of a null one from the VBC path. This leads to an inversion of the dominant path: it is now the indirect path **B2** which dominates the direct path **B1** in open ear. As the indirect path **B2**, which dominates the direct path **B1**, appears to be decreased by occlusion, the subjective OE is negative. Although the objective VABC OE has not been measured, we can deduce that it should be negative since the indirect path **B2a** is decreased by occlusion.

For the voice conduction path, the objective voice OE is negative since the indirect paths **A2a** and **B2a** appear to be decreased by occlusion. This is also justified by the fact the voice OE is a combination of the VBC one and the VABC one which are both negative.

1.7 Conclusions and Recommendations

In this article, several OE measurements are analyzed in order to characterize the "hollow voice" OE that creates a discomfort when people wear hearing aids or hearing protectors. A diagram of the internal sound path components involved in the perception of one's own voice is proposed in order to identify the different conduction paths of the internal voice. This internal voice path is subdivided into the VBC path due to a structure-borne source and the VABC path due to an airborne source. Both of these paths are subdivided into a direct path to the inner ear and an indirect one through the ear canal. The VBC path is characterized using literature results on objective and subjective BC OE measurements. For the VABC path, as no results were available in the literature, a new kind of subjective OE measurement is proposed and experimented in this article: a sound source (speaker) is placed at the input of the mouth. The internal voice path is described using the aforementioned VBC and VABC paths characterization and the literature results on objective voice OE measurements.

At low frequencies, in the case of the occluded ear, the measurements performed on the VBC and the VABC paths have shown that the indirect path through the ear canal to the inner ear dominates the direct one. It has been concluded from these observations that this is also true for the internal voice path which is the combination of these two paths as mentioned previously. Hence, the VBC path, the VABC path and the internal voice conduction path have identical behaviors. In the case of the open ear, the measurements used in this paper did not permit to draw conclusions. It has been found that the objective and subjective VBC OE, the subjective VABC OE and the objective internal voice OE are positive. This means that the energy transmitted by the indirect path is increased by the occlusion. Hence, we have been able to conclude that the objective VABC OE and the subjective internal voice OE should also be positive.

At high frequencies, the behaviors of the VBC, the VABC and the internal voice conduction path differ. In the case of the occluded ear, the direct path dominates the indirect one for the VBC path, whereas it is the inverse for the VABC path, hence no conclusion could be drawn for their combination, i.e. for the internal voice conduction path. In the case of the open ear, the dominating path has been identified for the VBC path only: the direct path dominates the indirect one. The objective OE is found to be negative in the three cases: from the literature, the objective VBC OE and the objective voice OE are negative, which leads to the conclusion that the objective VABC OE should also be negative. The subjective VBC OE is about zero, the subjective VABC OE is negative and the subjective internal voice OE is unknown.

Additional experiments such as objective OE measurement for the VABC path and more generally, non threshold subjective measurement method, need to be developed in order to further characterize the internal voice OE as well as its associated dominating path (direct versus indirect) in open and occluded ears.

Acknowledgements

The authors wish to express their appreciation to SONOMAX HEARING HEALTHCARE INC. (Montréal, Canada) and to NSERC (Natural Sciences and Engineering Research Council of Canada) for their support in this project.

References

E. Bárány, “A contribution to the physiology of bone conduction”, Acta Oto-Laryngol. **26**, 1–223 (1938).

G. v. Békésy, “The structure of the middle ear and the hearing of one’s own voice by bone conduction”, J. Acoust. Soc. Am. **21**, 217–232 (1949).

G. v. Békésy, “Bone conduction”, in *Experiments in hearing*, edited by E. G. Wever (McGraw Hill, 1960), 127–203.

E. H. Berger and J. E. Kerivan, “Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold”, J. Acoust. Soc. Am. **74**, 81–94 (1983).

E. H. Berger, “Hearing protection devices”, in “Noise & Hearing Conservation Manual”, edited by E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (American Industrial Hygiene Association, 1986), fourth edition, pp. 319–382.

H. Dillon, “Hearing aid earmolds, earshells and coupling systems”, in “Hearing aids” (Thieme, New York, 2000), pp. 117–158.

R. H. Ehmer, “Masking patterns of tones”, J. Acoust. Soc. Am. **31**, 1115–1120 (1959).

S. A. Gelfand, “Hearing : an introduction to psychological and physiological acoustics” (Marcel Dekker, New York, 2004), 4th edition.

M. O. Hansen, “Occlusion effects: Part 1: Hearing aid users experiences of the occlusion effect compared to the real ear sound level”, Technical Report 71, Technical University of Denmark (1997).

M. O. Hansen, “Occlusion effects: Part 2: A study of the occlusion effect mechanism and the influence of the earmould properties”, Technical Report 73, Technical University of Denmark (1998).

P. Howell, “Auditory feedback of the voice in singing”, in *Musical structure and cognition*, edited by P. Howell, I. Cross, and R. West (Academic Press, London, 1985), pp. 259–286.

Industrial Noise Laboratory, “Test report hearing protector noise attenuation - ANSI S12.6 - 1997 (b) - reat - subject fit”, Technical Report, Federal University of Santa Catarina (2007).

ISO 389-7, (1996). “Acoustics - reference zero for the calibration of audiometric equipment - part 7: Reference threshold of hearing under free-fied and diffuse-field listening conditions”.

ISO 4869-1, (1990). “Acoustics - hearing protectors - part 1: Subjective method for the measurement of sound attenuation”.

ISO 8253-1, (1989). “Acoustics - audiometric test methods - part 1: Basic pure tone air and bone conduction threshold audiometry”.

ISO 8253-2, (1992). “Acoustics - audiometric test methods - part 2: Sound field audiometry with pure tone and narrow-band test signals”.

M. C. Killion, “The hollow voice occlusion effect”, in *Proceedings of 13th Danavox Symposium*, volume 3, pp. 231–241 (1988).

P. Lundh, “Sound pressure in the ear with vented and unvented earmould”, Technical Report 28-8-1, Oticon Electronics A/S (1986).

S. Reinfeldt, S. Stenfelt, T. Good, and B. Hakansson, “Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations”, J. Acoust. Soc. Am. **121**, 1576–1587 (2007).

A. M. J. Small and R. S. Gales, “Hearing characteristics”, in *Handbook of acoustical measurements and noise control*, edited by C. M. Harris (Acoustical Society of America, Woodbury, NY, 1998), third edition, pp. 17.1–17.25 .

S. Stenfelt, N. Hato, and R. L. Goode, “Factors contributing to bone conduction: The middle ear”, J. Acoust. Soc. Am. **111**, 947–959 (2002).

S. Stenfelt, T. Wild, N. Hato, and R. L. Goode, “Factors contributing to bone conduction: the outer ear”, J. Acoust. Soc. Am. **113**, 902–913 (2003).

S. Stenfelt and S. Reinfeldt, “A model of the occlusion effect with bone-conducted stimulation”, Int. J. Audiol. **46**, 595–608 (2007).

J. Tonndorf, E. C. Greenfield, and R. S. Kaufman, “The occlusion of the external ear canal: its effect upon bone conduction in cats”, Acta Oto-Laryngol. **61**, 80–104 (1966).

J. Tonndorf, “Bone conduction”, in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic Press, New York, 1972), volume 2, pp. 195–237.

J. Voix, “Mise au point d’un bouchon d’oreille “intelligent” (Development of a “smart” earplug)”, Ph.d. thesis, École de Technologie Supérieure, Montréal, Canada (2006).

J. Voix and F. Laville, “The objective measurement of individual earplug field performance”, J. Acoust. Soc. Am. **125**, 3722-3732 (2009).

V. H. Wimmer, “The occlusion effect from earmoulds”, Hearing Instruments **37**, 19, 57–58 (1986).

CHAPITRE 2

ARTICLE #2

"WAVELET SPEECH ENHANCEMENT FOR INDUSTRIAL NOISE ENVIRONMENTS"

Cécile Le Cocq, Christian Gargour, Frédéric Laville,

École de technologie supérieure, Université du Québec,

Montréal (Québec), Canada, H3C 1K3

This article has been submitted to Speech Communication 20 January 2009.

Résumé

Dans un milieu industriel bruité, les travailleurs doivent souvent porter des protecteurs auditifs afin de préserver leur audition. Toutefois ces protecteurs et les signaux sonores de fort niveau diminuent leurs capacités à communiquer en altérant l'intelligibilité du signal de parole. Dans cet article, le débruitage de la parole est utilisé pour parer à cet inconvénient. Un grand nombre de méthodes classiques de débruitage par seuillage d'ondelettes sont testées sur plusieurs signaux de parole qui sont altérés par l'addition de bruits industriels selon une large gamme de rapports signal à bruit. Quatre critères de sélection sont utilisés afin de quantifier les résultats obtenus en terme de performance de débruitage. Une base de données de grande taille est ainsi constituée. Son analyse a nécessité la création d'un algorithme de sélection. Pour chacun des critères de sélection, cet algorithme calcule l'efficacité relative de chacune des méthodes de débruitage considérées. Cet algorithme est utilisé plusieurs fois dans la méthodologie générale afin de déterminer la ou les méthodes qui permettent d'obtenir les meilleurs résultats pour tous les cas testés.

Abstract

In industrial noise environments, workers often have to wear hearing protectors inorder to prevent hearing loss. However, hearing protectors and loud signals can interfere with speech intel-

ligibility. In this paper speech denoising is used to overcome this problem. A large number of speech denoising methods based on classical wavelet thresholding are tested on several speech signals which have been altered by the addition of industrial noises according to a broad range of signal to noise ratios. Four selection criteria are used to quantify the denoising performance results. A large database has been generated, requiring the creation of a selection algorithm. This algorithm determines the relative efficiency of each of the techniques under consideration for a given selection criterion. This algorithm is used several times in the general methodology to determine which specific methods yield the best results for all the tested cases.

2.1 Introduction

Workers in industrial noise environments are exposed to noise levels that could cause severe hearing damage. In 1995, the number of individuals suffering from disabling hearing difficulties was estimated to be 120 million worldwide (WHO, 1999, Ch. 3). To minimize the number of new cases of deafness, the World Health Organization (WHO) issued recommendations, some of which have been adopted in their entirety or in part in the legislature (WHO, 2001, Ch 4.). The first recommendation consists in limiting the equivalent noise exposure level over 8 hours per day to 85 dB(A). Wearing hearing protectors is an effective and low cost way to meet this recommendation. Most hearing protectors do not allow the user to differentiate useful signals such as speech or warning sounds from harmful noises; consequently, workers wearing hearing protectors find themselves in a distorted low-level sound environment which no longer enables them to understand one another or to perceive danger (Berger et al., 2000, Ch. 10). This communication problem could be solved with hearing protectors in which a microphone collects sound outside the protector, a digital signal processor denoises speech and warning signals and a speaker emits the denoised signal sound under the protector. This solution requires addressing the issue of speech enhancement in industrial noise environments; this is the object of the research presented in this paper.

Several speech denoising techniques already exist (e.g. Ephraim, 1992; Gustafsson et al., 2001; Haykin, 2002); the present study focuses on speech denoising methods by classical wavelet thresholding, the basis of this method was developed in the 1990s, by Donoho and Johnstone in their pioneering work (Donoho and Johnstone, 1994). This technique, which makes use of the wavelet transforms, consists in shrinking the wavelet coefficients below a certain threshold in order to remove noise and enhance speech. They proposed several thresholding rules and different ways of computing the threshold, notably soft and hard thresholding rules, universal and SURE (Stein unbiased risk estimator) thresholds (Donoho and Johnstone, 1995). The corresponding algorithms were originally designed for speech signals corrupted by white Gaussian noise. Johnstone and Silverman (Johnstone and Silverman, 1997) generalized these methods by adapting the threshold calculation to coloured noises. Following these developments, several variations of these algorithms were proposed, mainly by modifying the thresholding rule being applied. Sheikhzadeh presented the μ-law thresholding rule (Sheikhzadeh and Abutalebi, 2001), which fits between the hard and the soft thresholding rule.

A relatively small number of complete comparative studies on these speech denoising methods by classical wavelet thresholding can be found to date in the literature. The examples cited below present some of the more extensive comparative works undertaken in this field. In 2001, Antoniadis, Bigot and Sapatinas (Antoniadis et al., 2001) compared classical and bayesian wavelet denoising methods on a bank of theoretical mathematical signals altered by white Gaussian noise with different signal to noise ratios (SNRs). More recently, in 2003, Fodor and Kamath (Fodor and Chandrika, 2003) compared classical wavelet denoising methods applied to images altered by white Gaussian noise with more traditional imagery techniques using spacial filters. Finally, in 2006, Ayat, Manzuri-Shalmani and Dianat (Ayat et al., 2006) experimented with classical wavelet denoising methods on speech signals altered by white Gaussian noise.

Unlike the studies mentioned above (Antoniadis et al., 2001; Ayat et al., 2006; Fodor and Chandrika, 2003), the speech denoising is performed in this paper in the industrial noise environments. Our purpose is to find an algorithm which could be adequately used for hearing protectors. An industrial noise environment is defined, in the context of our work, as an environment in which workers are exposed to noises produced by machineries or any other industrial equipment. In this kind of environment, the SNRs extend on a wide range of values.

Hence the workers could have various levels of difficulties to understand speech depending upon their specific environment such as the distance at which they are from the noise and the sound sources. In this paper, the performance of certain classical wavelet denoising methods is evaluated in an industrial noise environment. While studies in the literature often use white Gaussian noise (Antoniadis et al., 2001; Ayat et al., 2006; Fodor and Chandrika, 2003), the speech signals considered in this study are altered by industrial noises recorded at factory workstations. Also, our study will cover a wide range of SNRs spanning from -20 dB to +20 dB, while the SNRs studied in the literature are rarely lower than -10 dB (Ayat et al., 2006; Fodor and Chandrika, 2003). In order to quantify the performance of the denoising methods considered, four selection criteria among the most commonly used (Hansen and Pellom, 1998; Wang et al., 1992) are considered namely: the global and segmented SNRs, the mean square error and the Itakura-Saito distortion measure. To ensure unbiased results given the interdependence of certain parameters, a global study is conducted to compare all the methods without having to set some of these parameters to specified values as it is often done in the literature (Ayat et al., 2006; Fodor and Chandrika, 2003). A selection algorithm has been developed to perform this global study.

This paper is organized as follows. Section 2.2 presents the denoising methods by classical wavelet thresholding theory with its various parameters: thresholding rule, threshold expression and noise estimate expression. In section 2.3, the signals, methods and selection criteria are presented. Section 2.4 explains the general methodology used to select the denoising methods. In section 2.5, the results are presented. Section 2.6 discusses both the approach and results. Lastly, section 2.7 presents the conclusions and recommandations.

2.2 Wavelet thresholding theory

If s denotes a pure speech signal and w a noise, the noised speech signal x can be defined as the sum of these two signals:

$$x = s + w \tag{2.1}$$

Denoising by wavelets consists in thresholding the wavelet coefficients of the noised signal. By applying the wavelet transform to Eq. (2.1) we obtain the following equation, where the capital letter represents the wavelet transform of the signal represented by the corresponding lower case one:

$$X = S + W \tag{2.2}$$

In this study, the considered algorithm is meant to be used for hearing protectors. Several types of wavelet transforms do exist. The wavelet transform to be utilized should meet two criteria: its implementation should be possible using a fast algorithm and it should permit perfect reconstruction of the signal. The dyadic wavelet transform meets these two criteria, as long as the analysis wavelet type is properly chosen.

Denoising by wavelet thresholding consists in applying a thresholding rule THR with a threshold T to the wavelet coefficients of the noised signal X to obtain the wavelet coefficients of the denoised signal $\tilde{S}$, which is an estimate of the wavelet coefficients of the pure speech signal S:

$$\tilde{S} = \mathrm{THR}\left(X, T\right) \tag{2.3}$$

The risk $r_{\mathrm{THR}}\left(s, T\right)$ associated to a thresholding rule THR with a given threshold T is defined by:

$$r_{\mathrm{THR}}\left(s, T\right) \;=\; \mathrm{E}\left\{\left\|S - \tilde{S}\right\|^2\right\} \tag{2.4}$$

The following sections present the three wavelet denoising parameters: the thresholding rule, the threshold expression and the noise estimate expression.

2.2.1 Thresholding rules

The first two thresholding rules have been formulated by Donoho and Johnstone (Donoho and Johnstone, 1994). they are the soft (THR_s) and the hard (THR_h) thresholding rules where:

$$\mathrm{THR}_s\left(X,T\right) = \begin{cases} \mathrm{sgn}(X)\left(|X|-T\right) & ,|X|\geq T \\ 0 & ,|X|<T \end{cases} \tag{2.5}$$

$$\mathrm{THR}_h\left(X,T\right) = \begin{cases} X & ,|X|\geq T \\ 0 & ,|X|<T \end{cases} \tag{2.6}$$

Sheikhzadeh (Sheikhzadeh and Abutalebi, 2001) has proposed the μ-law thresholding rule (THR_μ):

$$\mathrm{THR}_\mu(X,T) = \begin{cases} X & ,|X|\geq T \\ T\left(\frac{1}{\mu}\left[(1+\mu)^{|X/T|}-1\right]\mathrm{sgn}(X)\right), & |X|<T \end{cases} \tag{2.7}$$

This rule is intermediate between the hard and the soft thresholding rules: as the value of the parameter μ increases, the μ-law thresholding rule tends toward the hard thresholding one. It has three advantages over the hard and soft thresholding rules: first, it is a continuous function, secondly, no coefficient is set to zero and thirdly the coefficients greater than the threshold are not modified, thus the speech signal intelligibility is less affected by it than by the hard or soft thresholding rules (Nordström et al., 1999; Sheikhzadeh and Abutalebi, 2001).

2.2.2 Threshold expressions

The threshold expressions: universal, SURE and hybrid SURE, under consideration in this paper are presented in the next three sections. The fourth section presents the risk estimator expressions associated with the thresholding rules.

2.2.2.1 Universal threshold

The threshold value T can be determined in different ways. Donoho and Johnstone (Donoho and Johnstone, 1994) introduced the universal threshold T_u, which they defined for white Gaussian noise. Johnstone and Silverman (Johnstone and Silverman, 1997) adapted this threshold for coloured noises $T_u(j)$: the threshold is determined on each analysis level j of the wavelet transform instead of being determined globally:

$$T_u = \tilde{\sigma} \sqrt{2 \log N} \tag{2.8}$$

$$T_u(j) = \tilde{\sigma}_j \sqrt{2 \log N} \tag{2.9}$$

$\tilde{\sigma}$ and $\tilde{\sigma}_j$ are estimates of the standard deviation of the noise and will be presented in the following section (see Section 2.2.3) and N is the number of signal samples that are considered.

2.2.2.2 SURE threshold

To lower the risk associated to a given thresholding rule with the universal threshold, especially for low noise levels, Donoho and Johnstone (Donoho and Johnstone, 1995) proposed the SURE (Stein unbiased risk estimator) threshold T_{SURE}. As with the universal threshold, it was mainly meant for white Gaussian noise before being generalized by Johnstone and Silverman (Johnstone and Silverman, 1997) to coloured noises:

$$T_{\mathrm{SURE}} = \tilde{\sigma}\ \mathrm{SURE}(X/\tilde{\sigma}) \tag{2.10}$$

$$T_{\mathrm{SURE}}(j) = \tilde{\sigma}_j\ \mathrm{SURE}(X_j/\tilde{\sigma}_j) \tag{2.11}$$

$\mathrm{SURE}(X)$ denotes the threshold value T for which the estimator $\tilde{r}_{\mathrm{THR}}(s,T)$ of the risk associated to the thresholding rule THR is minimal:

$$\mathrm{SURE}(X) = \arg \min_{0 \leq T} \tilde{r}_{\mathrm{THR}}(s,T) \tag{2.12}$$

2.2.2.3 Hybrid SURE threshold

Donoho and Jonhstone (Donoho and Johnstone, 1995) brought to evidence that the SURE threshold tends to be too low when the noise level is very high. To mitigate this inconvenience, they proposed the hybrid SURE threshold $T_{\mathrm{SURE\,hybrid}}$, which replaces the SURE threshold by the universal threshold when the noise level is high:

$$T_{\mathrm{SURE\,hybrid}} = \begin{cases} T_{\mathrm{SURE}} & , \|X\|^2 - N\sigma^2 \leq \epsilon_N \\ T_u & , \|X\|^2 - N\sigma^2 > \epsilon_N \end{cases} \tag{2.13}$$

$$\text{with} \quad \epsilon_N = \sigma^2 \, N^{1/2} \, (\log N)^{3/2} \tag{2.14}$$

2.2.2.4 Risk estimator expression

In Eq. (2.12), an estimator of the risk associated to the thresholding rule THR is used since the true risk associated to a given thresholding rule expressed by Eq. (2.4) cannot be determined when the characteristics of the pure speech signal S are unknown. The risk estimator expressions for soft $\tilde{r}_{\mathrm{THR}_s}(s,T)$ (Donoho and Johnstone, 1995) and hard $\tilde{r}_{\mathrm{THR}_h}(s,T)$ (Krim et al., 1999) thresholding rules are:

$$\tilde{r}_{\mathrm{THR}_s}(s,T) = N - 2\,\#\left\{n : |X[n]| < T\right\} + \sum_{n=1}^{N} \left[\min\left(|X[n]|\,, T\right)\right]^2 \tag{2.15}$$

$$\tilde{r}_{\mathrm{THR}_h}(s,T) = N - 2\,\#\left\{n : |X[n]| < T\right\} + \sum_{|X[n]|<T} |X[n]|^2 \tag{2.16}$$

in which $\#\left\{n : |X[n]| < T\right\}$ denotes the number of different values for n that verify the inequality $|X[n]| < T$.

The risk estimator associated to soft thresholding rule is unbiased: $\mathrm{E}\left\{\tilde{r}_{\mathrm{THR}_s}(s,T)\right\} = r_{\mathrm{THR}_s}(s,T)$ (Donoho and Johnstone, 1995). However, the estimator associated to hard thresholding rule is biased and its bias can be expressed as follows (Krim et al., 1999):

$$\begin{aligned} \mathrm{Bias}_h &= r_{\mathrm{THR}_h}(s,T) - \mathrm{E}\left\{\tilde{r}_{\mathrm{THR}_h}(s,T)\right\} \\ &= 2\,T\,\sigma^2 \sum_{n=1}^{N} \left[\phi\left(T - S[n]\right) + \phi\left(T + S[n]\right)\right] \end{aligned} \tag{2.17}$$

in which $\phi(u) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left(u^2/2\sigma^2\right)}$

For the μ-law thresholding rule, based on the definition, we have formulated the true associated risk as:

$$r_{\mathrm{THR}_\mu}(s,T) = \sum_{n=1}^{N} \mathrm{E}\left\{ \left| S[n] - T\left(\frac{1}{\mu}\left[(1+\mu)^{|X[n]/T|} - 1\right]\mathrm{sgn}(X[n])\right)\right|^2 \right\} \tag{2.18}$$

However, a risk estimator associated to this rule is more complex to obtain because of the presence of the term $(1+\mu)^{|X[n]/T|}$ in Eq. (2.18).

2.2.3 Noise estimate expressions

Three main different approaches can be used to estimate the standard deviation of the noise based on the noised signal.

In the first approach the noise added to the speech signal has a unit variance ($\tilde{\sigma} = 1$). This is a common procedure.

In the second approach, Donoho and Johnstone (Donoho and Johnstone, 1994) proposed the estimation of the white Gaussian noise standard deviation using the median absolute deviation (MAD) calculated on the first analysis level of the wavelet transform of the noised signal.

$$\tilde{\sigma} = \frac{\mathrm{MAD}(X_1)}{0.6745} \tag{2.19}$$

with $\mathrm{MAD}\,(X_1) = \mathrm{median}\,(|X_1|)$

In the third approach, Johnstone and Silverman (Johnstone and Silverman, 1997) adapted Donoho and Johnstone's estimator to coloured noises by determining an estimate for each analysis level j of the wavelet transform of the noised signal.

$$\tilde{\sigma}_j = \frac{\mathrm{MAD}(X_j)}{0.6745} \tag{2.20}$$

with $\mathrm{MAD}\,(X_j) = \mathrm{median}\,(|X_j|)$

2.3 Presentation of the studied cases

The performance of wavelets thresholding methods presented (see Section 2.2) is studied for experimental signals formed by speech signals and industrial noises. The quality of the denoising obtained is quantified using four selection criteria. A database of 42 508 800 values of the different criteria is generated, its composition is given in Table 2.1. Firstly, the experimental signals used in simulations are presented (see Section 2.3.1): 8 200 signals are tested. Secondly, the parameters of wavelet denoising methods tested are exposed in details (see Section 2.3.2): 1 296 denoising methods are tested. Thirdly, the four selection criteria used to quantify the quality of the denoising achieved are explained (see Section 2.3.3).

2.3.1 The signals

The twenty speech signals used come from the TIMIT database (Garofolo et al., 1993). Ten English phrases spoken by a woman and the same ten phrases spoken by a man were selected.

For this study, an industrial noise is considered to be any noise to which a worker is exposed in an industrial workplace. Ten industrial noises are used. Two are extracted from the NOISEX database (NOISEX , 1990), they are recorded in a car factory. The eight others are recorded in different locations at the NORANDA CCR copper refinery in Montreal (J. Voix, personal communication, 2000).The ten industrial noises tested could almost be considered as stationary since their frequency components vary only slightly over time. In addition to these recorded noises, white and pink Gaussian noises are tested to compare their behaviour to the industrial noises one and to identify which one of the two best simulates industrial noises in a speech denoising treatment.

Thus, two hundred (20×10) speech-noise pairs are formed. Every noised speech signal x is obtained by adding an industrial noise w to a speech signal s according to Eq. (2.21).

$$x = \frac{s}{\mathrm{std}(s)} 10^{\mathrm{SNR}/20} + \frac{w}{\mathrm{std}(w)} \tag{2.21}$$

with $\mathrm{std}(s) = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} (s[n] - \bar{s})^2}$ denoting the standard deviation of the signal s.

The two signals are first brought down to a unit variance. The signal to noise ratio SNR is chosen between -20 dB and 20 dB by 1 dB increment. A total of 41 SNRs are considered. Hence, a total of $20 \times 10 \times 41 = 8\ 200$ noised speech signals x are used to evaluate the denoising methods studied.

Table 2.1 Signals, methods and criteria

Index		Item	Type			Nb of items
1			**Signals**			
	a	Speech	Men	10 phrases		20
			Women	10 phrases		
	b	Noise	Noisex	car factory	2	10
			Noranda	copper refinery	8	
	c	SNR	$[-20:1:20]$ dB			41
Nb of signals						$20 \times 10 \times 41 = 8\,200$
2			**Methods**			
	a	Wavelet	Daubechies	1, 4, 8		8
			Symlets	4, 8		
			Coiflets	1, 2, 4		
	b	Nb of levels	6, 8, 10			3
	c	Denoising techniques parameters				
	α	Thresholding	hard		6	
			soft			
			μ-law	$\mu = 10^2,\ 10^3,\ 10^4,\ 10^5$		
	β	Threshold	universal		3	
			SURE			
			hybrid SURE			
	γ	Noise estimate	$\tilde{\sigma} = 1$		3	
			first level			
			each level			
	Nb of denoising techniques					$6 \times 3 \times 3 = 54$
Nb of methods						$8 \times 3 \times 54 = 1\,296$
3			**Criteria**			
	a	SNR	Global			2
			Segmental			
	b	MSE				1
	c	IS				1
Nb of criteria						$2+1+1 = 4$
Total nb of criteria values:						$8\,200 \times 1\,296 \times 4 = 42\,508\,800$

2.3.2 The methods

To achieve the inverse wavelet transform of the denoised signal, the analysis wavelet type chosen must ensure perfect reconstruction of the signal. Daubechies, Symlets, Coiflets and biorthogonal wavelets have this property. Unlike the other types, biorthogonal wavelets use two forms of wavelets: one for analysis and one for reconstruction. For a given order of filter, this family of wavelets uses twice the amount of memory as the other possible families and will therefore not be included in this study. We will consider only the Daubechies (db), Symlets (sym) and Coiflets (coif) wavelets. The choice of the wavelet order is also based on the amount of available memory: M order Daubechies and Symlets wavelets require 2M length filters while M order Coiflets wavelets require 6M length filters. The wavelets tested were chosen based on their current usage in the literature (Bui and Chen, 1998; Fan and Xia, 2001) and to limit the length of the filters: Daubechies wavelets order 1, 4 and 8; Symlets order 4 and 8; and Coiflets order 1, 2 and 4. To keep the algorithm calculation time within acceptable limits and still obtain sufficient accuracy, the closer number of analysis levels of the wavelet transform, which we have utilized, are 6, 8 and 10. Therefore, a choice of 8 analysis wavelet types applied to 3 values for the number of analysis levels are considered.

The wavelet thresholding algorithm (see Section 2.2) involves three parameters: the thresholding rule, the threshold expression and the noise estimate expression. The three thresholding rules presented (see Section 2.2.1) are considered here, i.e. soft, hard and μ-law thresholding rules. The third one is tested using four different μ parameter values (100, 1 000, 10 000 and 100 000). The four μ-law thresholding rules thus defined are distributed uniformly over a $\log$ scale (see Fig. 2.1) between the hard thresholding rule and a thresholding rule that would not modify the signal ($\mathrm{THR}(X, T) = X$). Therefore, six thresholding rules (soft thresholding, hard thresholding and four μ-law thresholdings) are considered for this study. Fig. 2.1 shows each of the thresholding rules considered. The three threshold expressions proposed (see Section 2.2.2) are considered: the univeral threshold, SURE threshold and hybrid SURE threshold. For the SURE and hybrid SURE thresholds, a thresholding rule risk estimator is necessary. For soft thesholding, the estimator is unbiased and could therefore be used to determine the threshold. For hard thresholding rule, the estimator is biased. Therefore, the SURE threshold is lower than the value it would attain with an unbiased estimator (Krim et al., 1999). The thresholding rule tends to “let through” more coefficients than necessary, therefore it “let through” more noise. Denoising results are therefore poorer than it could be expected. For

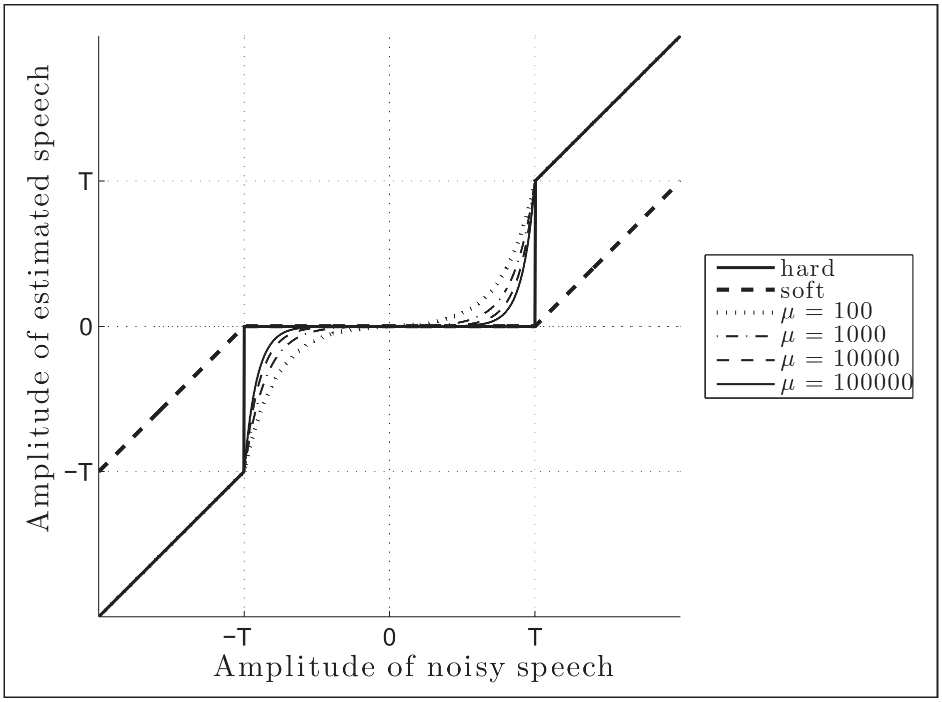

Figure 2.1 Thresholding rules.

μ-law thresholding rule, the risk estimator associated to the thresholding rule (see Eq. (2.18)) is not easily obtainable. Since we had no optimal estimator of the risk associated to hard and μ-law thresholding rules, we used the risk associated to the soft thresholding rule for these two thresholding rules. As for noise estimates, the three expressions presented (see Section 2.2.3) are used.

A total of $6 \times 3 \times 3 = 54$ denoising techniques are considered. These 54 denoising techniques are indexed in Table 2.2 for later identification. So a total of $8 \times 3 \times 54 = 1\ 296$ denoising methods are considered for this study.

Table 2.2 Denoising techniques

Index	Thresholding	Threshold	Noise estimate
1	hard	universal	$\tilde{\sigma} = 1$
2			first level
3			each level
4		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
5			first level
6			each level
7		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
8			first level
9			each level
10	soft	universal	$\tilde{\sigma} = 1$
11			first level
12			each level
13		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
14			first level
15			each level
16		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
17			first level
18			each level
19	μ-law $\mu = 100$	universal	$\tilde{\sigma} = 1$
20			first level
21			each level
22		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
23			first level
24			each level
25		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
26			first level
27			each level
28	μ-law $\mu = 1\,000$	universal	$\tilde{\sigma} = 1$
29			first level
30			each level
31		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
32			first level
33			each level
34		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
35			first level
36			each level
37	μ-law $\mu = 10\,000$	universal	$\tilde{\sigma} = 1$
38			first level
39			each level
40		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
41			first level
42			each level
43		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
44			first level
45			each level
46	μ-law $\mu = 100\,000$	universal	$\tilde{\sigma} = 1$
47			first level
48			each level
49		$\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
50			first level
51			each level
52		hybrid $\mathrm{SURE}_{\mathrm{soft}}$	$\tilde{\sigma} = 1$
53			first level
54			each level

2.3.3 The criteria

In order to quantify the objective performance of results obtained for each wavelet denoising method tested, four selection criteria of three different types were used (Hansen and Pellom, 1998; Wang et al., 1992):

a. Two criteria quantify the improvement in signal to noise ratio:

α. The global signal to noise ratio is defined as:

$$\mathrm{SNR_{glo}} \;=\; 10\ \log_{10}\left\{\frac{\mathrm{var}(s)}{\mathrm{var}(s-\tilde{s})}\right\} \tag{2.22}$$

with $\mathrm{var}(s) \;=\; \frac{1}{N-1}\sum\limits_{n=0}^{N-1}(s[n]-\bar{s})^2$ denoting the signal variance and $\bar{s} \;=\; \frac{1}{N}\sum\limits_{n=0}^{N-1}s[n]$ denoting the average signal.

It is one of the most popular measures for establishing signal denoising quality. It is the measure by which Ayat, Shalmani and Dianat (Ayat et al., 2006) evaluate the quality of their speech enhancement algorithm.

β. The segmental signal to noise ratio is defined as:

$$\mathrm{SNR_{seg}} = \frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\left\{\frac{\mathrm{var}(s_m)}{\mathrm{var}(s_m-\tilde{s}_m)}\right\} \tag{2.23}$$

with $\{s_m\}_{m=0\cdots M-1}$ the M frames of the signal s.

Derived from the global SNR, the segmental SNR is the average of the SNRs calculated for each signal frame. A speech signal is non-stationary, its frequency composition varies over time; therefore the segmental SNR provides a more accurate quality measure of the denoised speech signal than the global SNR (Hansen and Pellom, 1998).

b. One statistical criterion quantifies the errors introduced by denoising. It is the mean square error defined as:

$$\mathrm{MSE} \;=\; \frac{1}{N}\sum_{n=0}^{N-1}(s[n]-\tilde{s}[n])^2 \tag{2.24}$$

The MSE is commonly utilized as a comparison measure between the original signal and the denoised one. It has been used by Antoniadis, Bigot and Sapatinas (Antoniadis et al., 2001) and by Fodor and Kamath (Fodor and Chandrika, 2003).

c. One criterion quantifies the distortion of the speech signal: the Itakura-Saito distortion measure defined as:

$$\mathrm{IS} \quad = \quad \frac{1}{M} \sum_{m=0}^{M-1} \left\{ \left[\frac{G_{s_m}^2}{G_{\tilde{s}_m}^2} \right] \left[\frac{\vec{a}_{\tilde{s}_m} \mathbf{R}_{s_m} \vec{a}_{\tilde{s}_m}^T}{\vec{a}_{s_m} \mathbf{R}_{s_m} \vec{a}_{s_m}^T} \right] \; + \; \log \left(\frac{G_{\tilde{s}_m}^2}{G_{s_m}^2} \right) \; - \; 1 \right\} \tag{2.25}$$

in which $\mathbf{R}_{s_m}$ denotes the autocorrelation matrix of the frame s_m, $\vec{a}_{s_m}$ the linear prediction filter ceofficients of the frame s_m and $G_{s_m}^2$ defined by $G_{s_m}^2 = \mathbf{R}_{s_m}(1,:) \cdot \vec{a}_{s_m}^T$.

The IS distortion measure is seldom used in the literature in this context (Lu and Wang, 2007). It permits however an objective quantification of the denoised speech signal intelligibility (Hansen and Pellom, 1998).

2.4 Methods selection methodology

The database of 42 508 800 values obtained with the four selection criteria considered is very large and hence cannot be easily analyzed. Moreover, the various denoising method parameters are not all independent one from each other. The analysis wavelet type and the number of analysis levels are mutually independent and are also independent from the other parameters. On the other hand, the three denoising techniques parameters (the thresholding rule, the threshold expression and the noise estimate expression) are not mutually independent: the mathematical expression for the thresholding rule depends on the threshold expression which depends on the noise estimate expression (see Section 2.2). Therefore it has been chosen in this paper to study globally the influence of these three parameters on the denoising performance. The performance of the 54 denoising techniques should therefore be studied simultaneously, for each of the four selection criteria considered. For this purpose, we designed an appropriate algorithm to study the performance of these 54 denoising techniques. The general methodology used to process the database of 42 508 800 values consists of four steps, which are defined below and presented in Fig. 2.2.

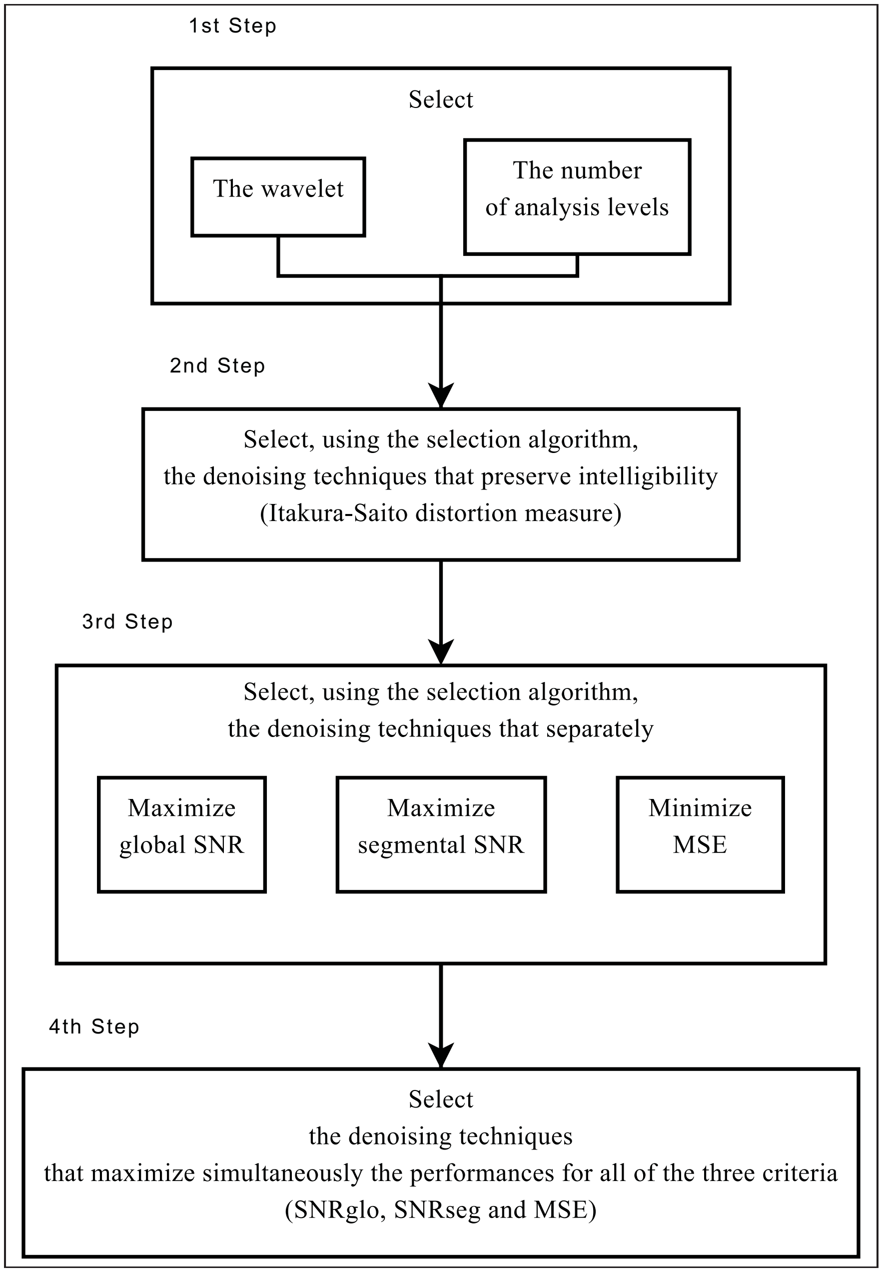

Figure 2.2 Main flowchart of the general methodology used for methods selection.

In the first step, the wavelet type and the number of analysis levels that will give the best results are determined by comparing the average of the best performance results for the various experimental signals according to each selection criteria. In the second step, the denoising techniques that preserve intelligibility are determined based on the IS distortion measure criterion using the selection algorithm. In the third step, the techniques that give the best results separately in terms of $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ or MSE are determined using the selection algorithm. In the fourth step, the techniques that give the best results simultaneously for all the three selection criteria, $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE are identified. These four steps are explained in the four following sections. The selection algorithm itself is presented in the fifth section.

2.4.1 First step: Selecting the adequate wavelet type and the number of analysis levels

The first step consists in applying the same algorithm twice to determine firstly the type of analysis wavelet and secondly, the number of analysis levels that will separately give the best performance according to the 4 selection criteria considered, namely $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$, MSE and IS. The used methodology is presented in Fig. 2.3 and explained below:

a. At this point, the entire database is considered.
b. The 42 508 800 values of the database are divided into 4 sets of 10 677 200 values. Each set contains values obtained for each of the four selection criteria ($\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$, MSE and IS).
c. Each of the 4 sets is divided into n sub-sets (8 for the wavelet types and 3 for the number of analysis levels) according to the possible options for the parameter considered. As an example, for wavelet types, a sub-set represents the set of results obtained during the use of Daubechies wavelets of order 1 (db1), another for db4, another for db8, another for Symlet wavelets of order 4 (sym4), etc.
d. Each of these $4 \times n$ sub-sets is divided into 41 groups according to 41 possible values of the SNR of the noised signal (see Table 2.1).
e. For each of these $4 \times n \times 41$ groups, the average for the 200 (20×10) speech-noise pairs for the best performance results (max or min) obtained is calculated.
f. Four figures (see for example Figs. 2.7 and 2.8 which will be presented in section 2.5.1), one per selection criterion, show these averages, according to the SNR of the noised signal for each of the n possible options for the parameter considered.
g. The option for the considered parameter (wavelet type or number of analysis levels) providing the best average of the best performance results for the entire set of SNR of the noised signal, according to the four selection criteria ($\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$, MSE and IS), is chosen.

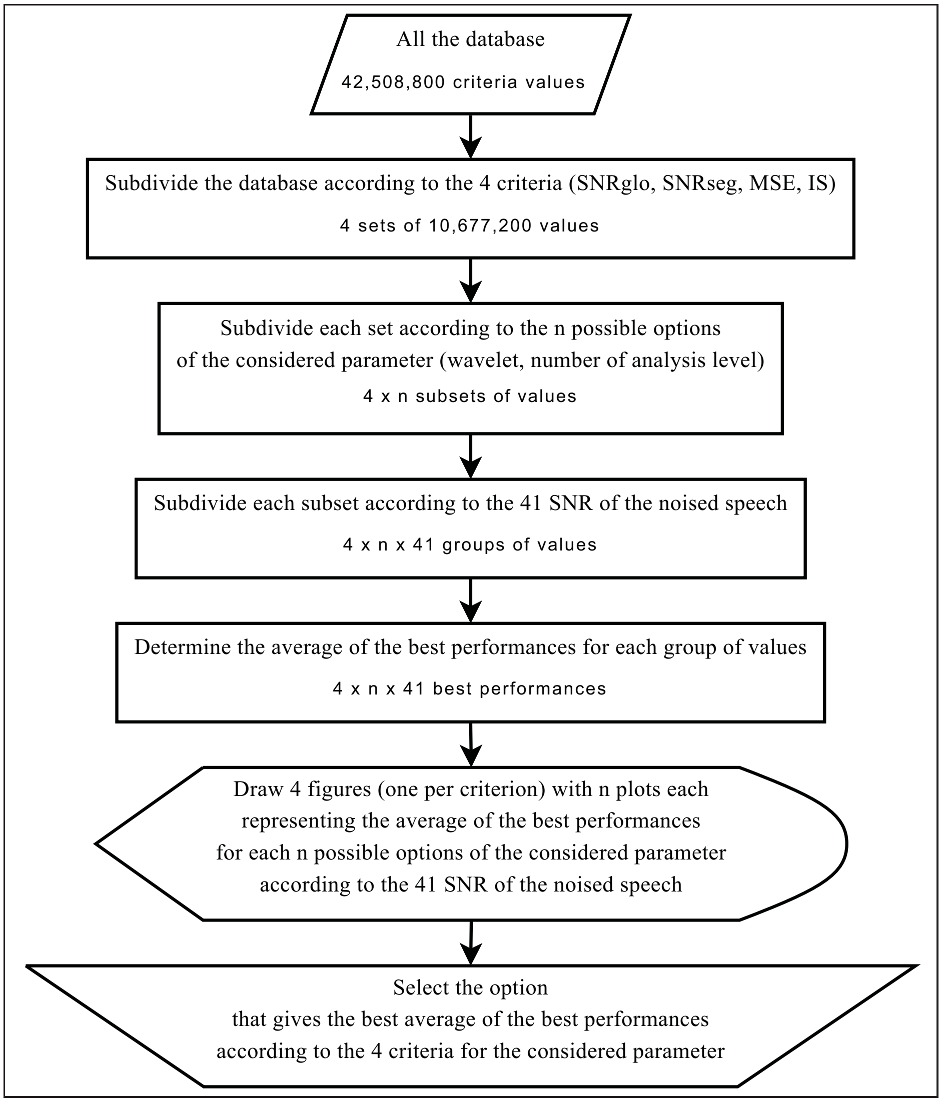

Figure 2.3 Flowchart of first step: Selecting the adequate wavelet type and the number of analysis levels.

At the end of this first step, the analysis wavelet type and the number of analysis levels, that enable us to obtain the best average for the best performance results according to the four selection criteria, $\mathrm{SNR}_{\mathrm{glo}}$, $\mathrm{SNR}_{\mathrm{seg}}$, MSE and IS, are determined.

2.4.2 Second step: Selecting the denoising techniques that preserve intelligibility

The second step consists in selecting the techniques that will preserve intelligibility of the denoised signal by using the IS distortion measure. The methodology is summarized in Fig. 2.4 and explained below:

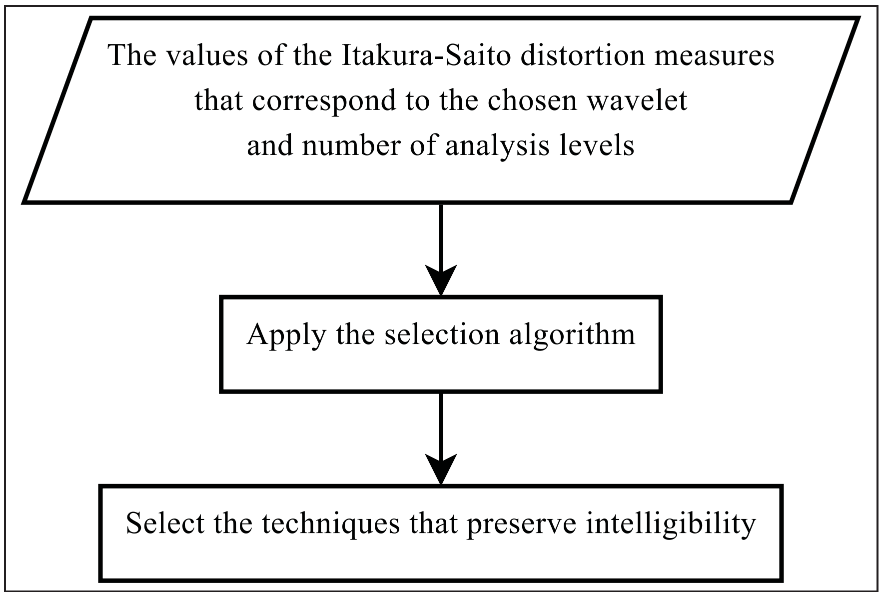

Figure 2.4 Flowchart of second step: Selecting the denoising techniques that preserve intelligibility.

a. The values considered here are the IS distortion measures that correspond to the choice made in step one in terms of the analysis wavelet type and of the number of analysis levels.
b. The selection algorithm, which will be explained in Section 2.4.5, is applied to the IS distortion measures considered in a.
c. The techniques that ensure a preset minimum level of intelligibility of the denoised signal are selected; their efficiency must not be zero (the "efficiency" is defined in this article as the percentage of speech-noise pairs for which the technique has been selected).

The set of parameters (thresholding rule, threshold expression and noise estimate expression) that will preserve intelligibility of the denoised signal are identified from the outcome of this second step.

2.4.3 Third step: Selecting the denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE

The third step consists in applying the same algorithm three times to determine the techniques that first maximize the $\mathrm{SNR_{glo}}$ criterion, second maximize the $\mathrm{SNR_{seg}}$ criterion and third minimize the MSE criterion. The methodology which is used and presented in Fig. 2.5 and explained below is applied on each of the three above mentioned criteria:

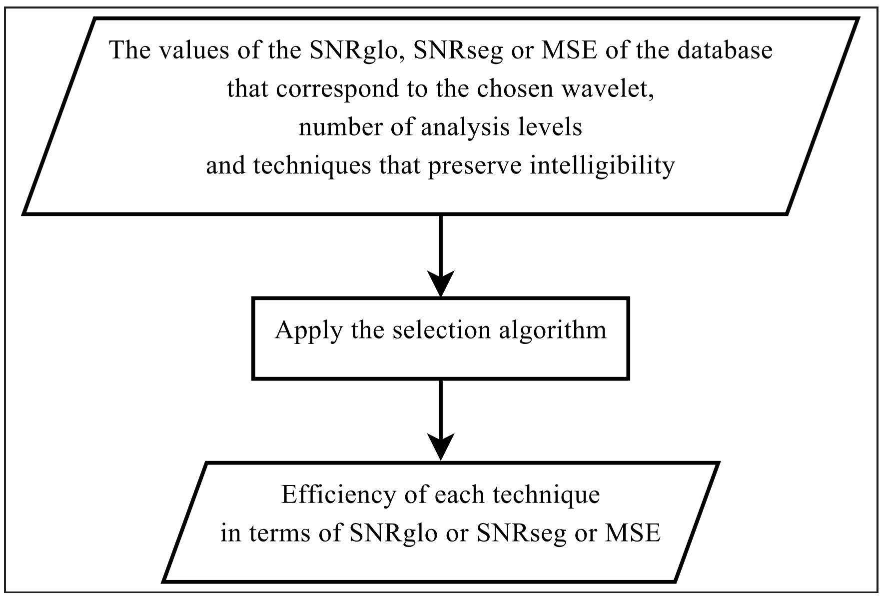

Figure 2.5 Flowchart of third step: Selecting the denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$**,** $\mathrm{SNR_{seg}}$ **and** MSE**.**

a. The values considered here are those of one of the three criteria $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE that correspond to the choices made in the first two steps in terms of analysis wavelet type, number of analysis levels and denoising techniques that preserve signal intelligibility.
b. The selection algorithm, which has already been used in the second step and which will be explained in Section 2.4.5, is applied to the values considered in a.
c. The efficiency for each of the techniques considered in terms of $\mathrm{SNR_{glo}}$ or $\mathrm{SNR_{seg}}$ or MSE, is obtained. For techniques that do not preserve signal intelligibility, the results are set to zero for all the three criteria $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE. The results obtained are presented

in a histogram (see for example Fig. 2.10 which will be presented is section 2.5.3), which represents the efficiency of each of the 54 denoising techniques considered.

2.4.4 Fourth step: Selecting the denoising techniques that simultaneously maximize performance for all the three criteria $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE

The fourth step consists in regrouping the results obtained in step three for all the three selection criteria, $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE. For each of these three selection criteria, a histogram (see for example Fig. 2.10 which will be presented is section 2.5.3), is plotted to show the efficiency of each of the 54 techniques considered. The comparison of these three histograms highlights the most efficient techniques.

2.4.5 Selection algorithm

Fig. 2.6 is a flowchart showing the selection algorithm. Its data input are the results obtained for a given criterion ($\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$, MSE and IS), for 200×41 signals (200 speech-noise pairs for which the signals are added according to 41 SNR) denoised using the techniques considered. Here are the different steps performed for each speech-noise pair:

a. Only the selected database values that correspond to one of the selection criteria (IS for the second step of the general methodology, $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ or MSE for its third step as shown in Fig. 2.2), are considered.
b. The set of values to process are divided into 200 sub-sets corresponding to the results obtained for each of the 200 speech-noise pairs.
c. A minimum-quality test is then performed. Each of the 200 sub-sets, i.e. each speech-noise pair, is tested to ensure that a minimum quality of denoising is obtained. For example, if the sub-set of values considered contains the gains in terms of $\mathrm{SNR_{glo}}$ for a given speech-noise pair, this sub-set is first divided into 41 groups according to the 41 possible values of the noised signal SNR. For each of these 41 groups, the maximum gain in terms of SNR is determined. The average of these 41 maximum gains is then calculated. This average of maximum gains must be greater or equal to the preset base value (1 dB for example)

to ensure that the speech-noise pair is valid. This base value is chosen depending on the type of criterion, the noised speech signals, the denoising techniques considered and the minimum quality one subsequently wishes to ensure. The study on the speech-noise pair is continued only if the minimum-quality test condition is satisfied. If not, the speech-noise pair considered cannot properly be denoised using the methods considered.

d. Each of the M sub-sets that meet the quality-test is divided into 41 groups according to the 41 possible values of the SNR of the noised signal.

e. For each of the $M \times 41$ groups, the denoising techniques that provide the best performance results are selected according to the following rule: the average of the performance should not be too far from the extremums obtained for the methods retained. Two parameters, chosen according to the type of criterion, the noised speech signals and the denoising techniques considered, enable us to quantify these deviations: an absolute deviation and a relative deviation between the maximum and minimum values. For example, if the group of values considered contains gains in terms of $\mathrm{SNR}_{\mathrm{glo}}$, the absolute deviation between the maximum and minimum values obtained for the retained denoising techniques must be lower than or equal to the preset maximum absolute deviation value (4 dB for example). The same procedure is applied for a preset maximum relative deviation (20% for example).

f. For every one of the M speech-noise pairs, the results obtained from the previous step for each of the 41 SNR are examined to identify and retain the techniques that yield the best performance results for the whole set of the 41 SNR values of the noised signal.

g. For each of the denoising techniques considered for the algorithm, the percentage of speech-noise pairs, for which this technique yields the best performance for all the 41 SNR of the noised signal, is calculated. This value is the “efficiency” defined is section 2.4.2.

The selection algorithm therefore enables us to determine the efficiency of each of the denoising techniques considered. The algorithm is blind: it has no information about the techniques it is testing and thus it cannot favour one over another.

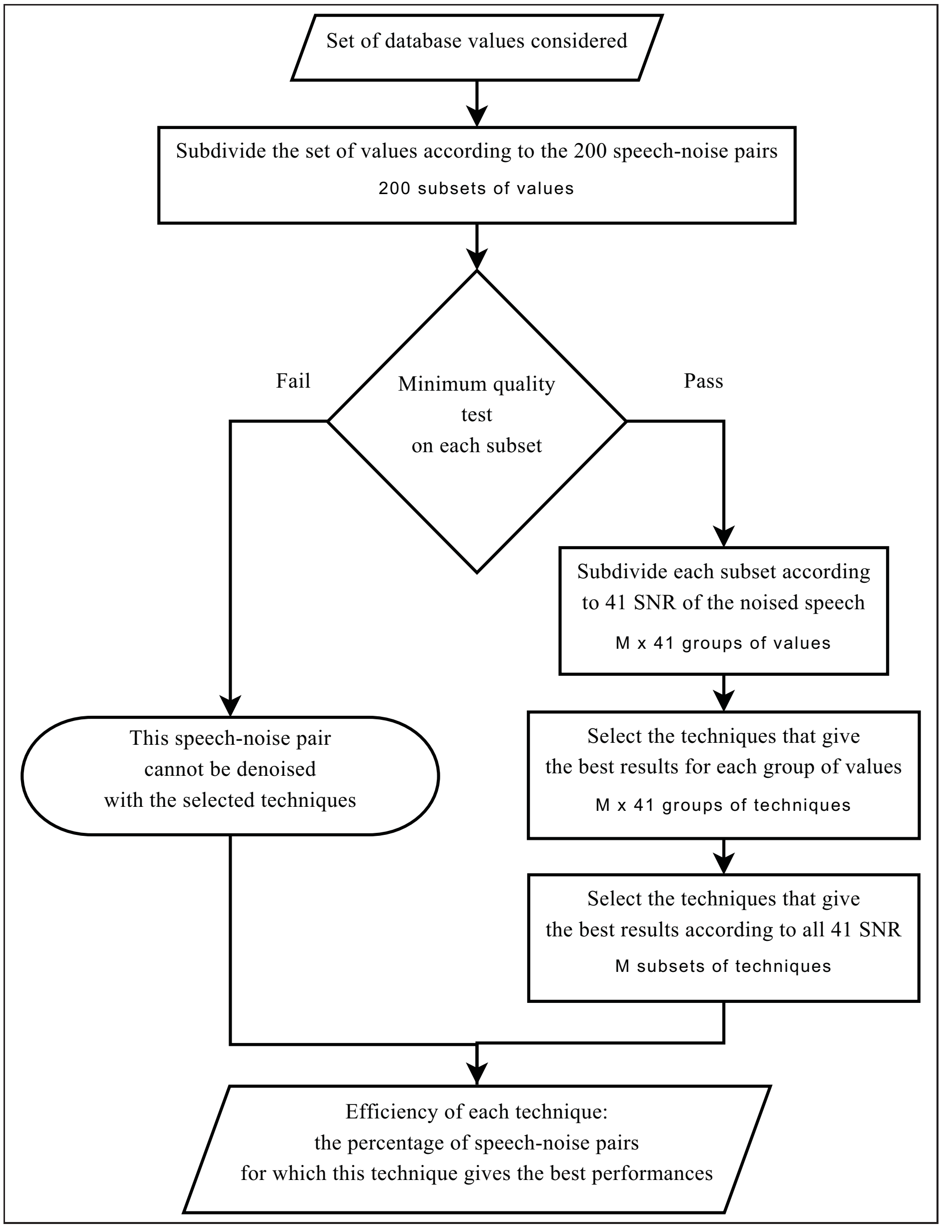

Figure 2.6 Flowchart of the selection algorithm.

2.5 Experimental results

2.5.1 First step: Choice of the analysis wavelet type and of the number of analysis levels

2.5.1.1 Choice of the analysis wavelet type

The results obtained for the choice of the analysis wavelet type in step 1 of the general methodology are shown in Fig. 2.7.

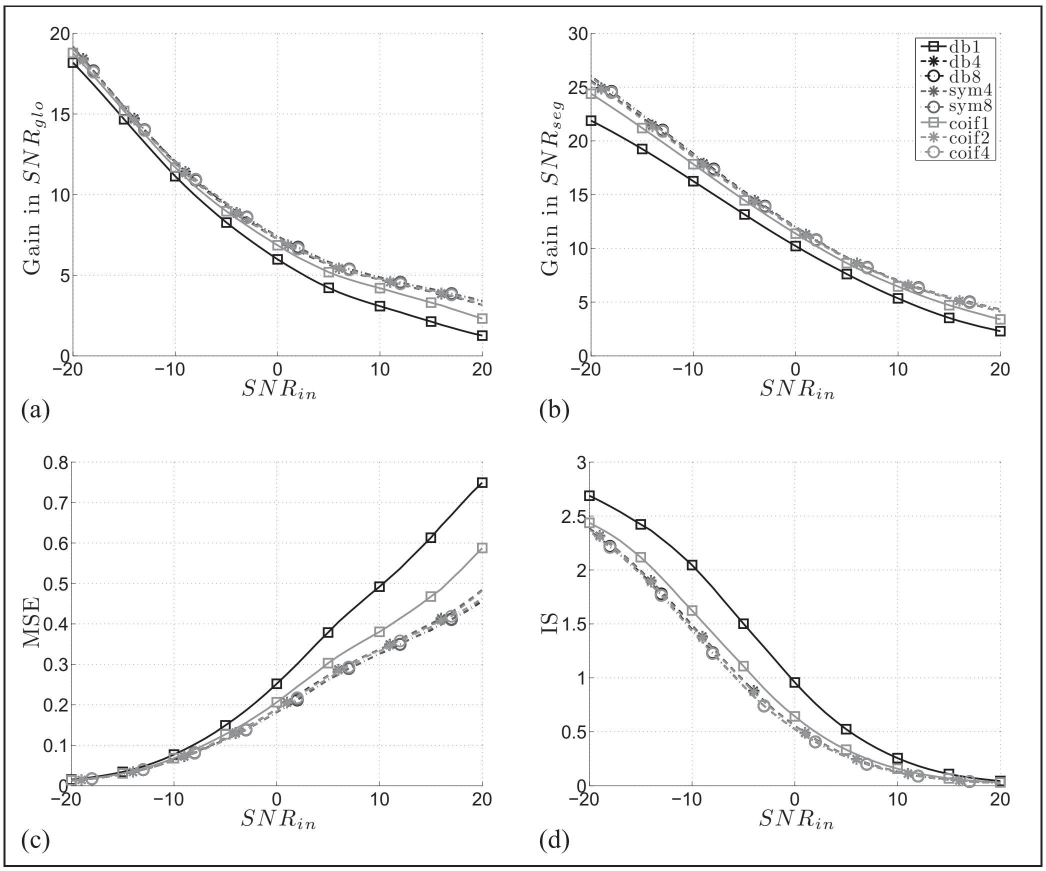

Figure 2.7 First step experimental results: Choice of the analysis wavelet type. (a) Gain in $\mathrm{SNR}_{\mathrm{glo}}$. (b) Gain in $\mathrm{SNR}_{\mathrm{seg}}$. (c) MSE. (d) IS.

If we consider the results in terms of the denoising gain of the global SNR, we notice that, except for the Daubechies wavelet of order 1, which does not perform as well as the other wavelets, especially when the SNR of the noised signal is positive, all the wavelets considered perform similarly over the entire range of SNRs of the noised signal ($\mathrm{SNR}_{\mathrm{in}}$). Therefore, among the tested analysis wavelet types, the one that gives the best performance with the smallest calculation time is the Daubechies wavelet of order 4. The results obtained for the three other selection criteria (segmental SNR, MSE and IS distortion measure) are similar.

2.5.1.2 Choice of the number of analysis levels

The results obtained for the choice of the number of analysis levels in step 1 of the general methodology are shown in Fig. 2.8.

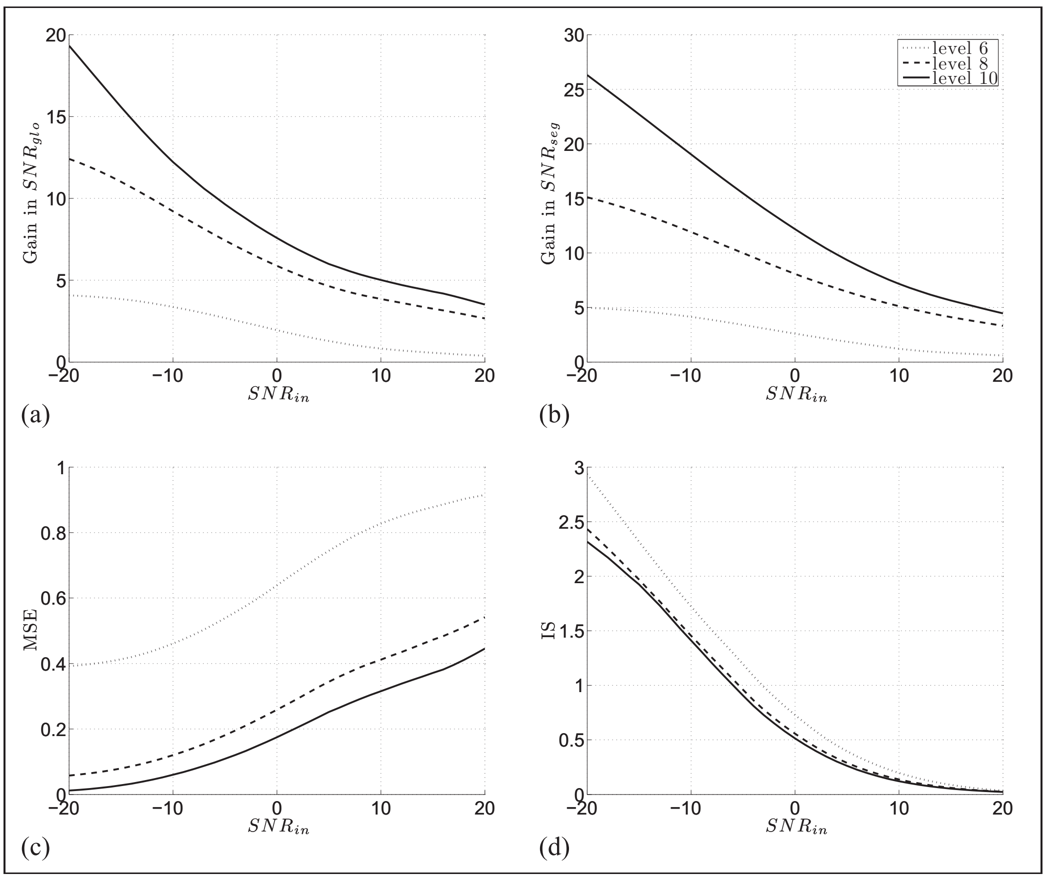

Figure 2.8 First step experimental results: Choice of the number of analysis levels. (a) Gain in $\mathrm{SNR}_{\mathrm{glo}}$. (b) Gain in $\mathrm{SNR}_{\mathrm{seg}}$. (c) MSE. (d) IS.

If we consider the results in terms of gain in segmental SNR, the more the SNR of the noised signal ($\mathrm{SNR}_{\mathrm{in}}$) is negative, the more the gain obtained by increasing the number of analysis levels of the wavelet transform is significant. Good results have been found to be obtained using ten analysis levels. The three other criteria (global SNR, MSE and IS distortion measure) lead to the same conclusion.

2.5.2 Second step: Selection of the denoising techniques that preserve intelligibility

The results obtained in step 2 of the general methodology for selecting the denoising techniques that preserve intelligibility by using the IS distortion measure are indicated in Fig. 2.9 . The efficiency of the 54 denoising techniques is shown. The histogram in Fig. 2.9 is presented with an ordinate in logarithmic scale in order to properly distinguish techniques with low denoising efficiency from those with zero efficiency.

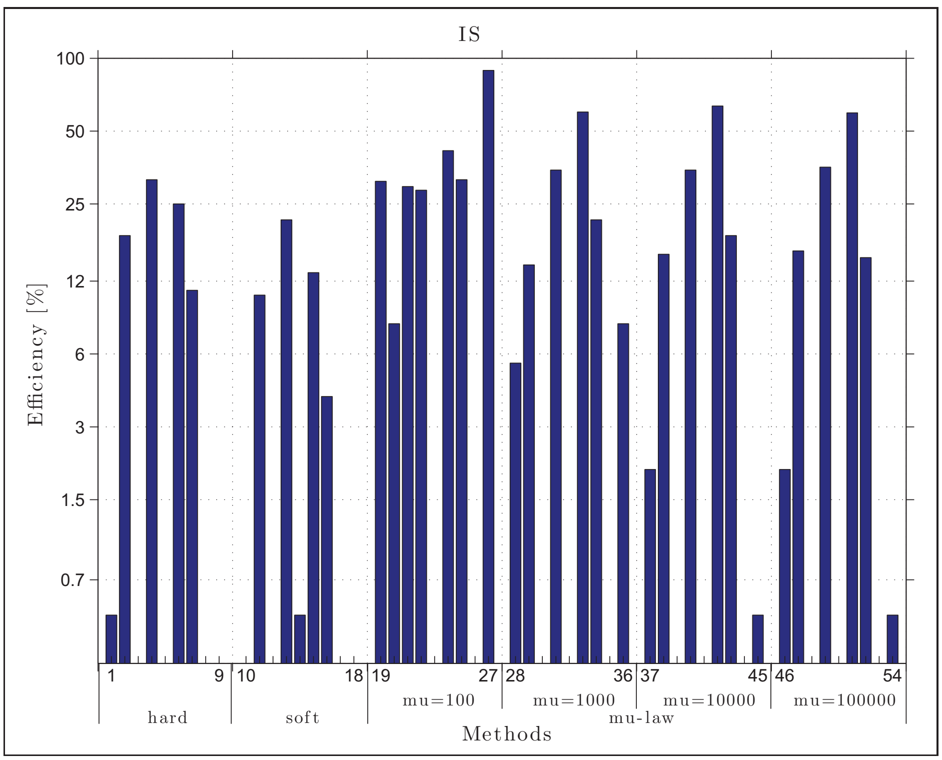

Figure 2.9 Second step experimental results: Selection of denoising techniques that preserve intelligibility.

Of the 54 denoising techniques tested, all those that have a nonzero efficiency, that is all those that preserve the intelligibility for at least one speech-noise pair are considered in step three of the general methodology.

2.5.3 Third step: Selection of the denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE

The results obtained in step 3 of the general methodology for selecting the denoising techniques that separately maximize the gain in global SNR, maximize the gain in segmental SNR and minimize the MSE are shown in Fig. 2.10.

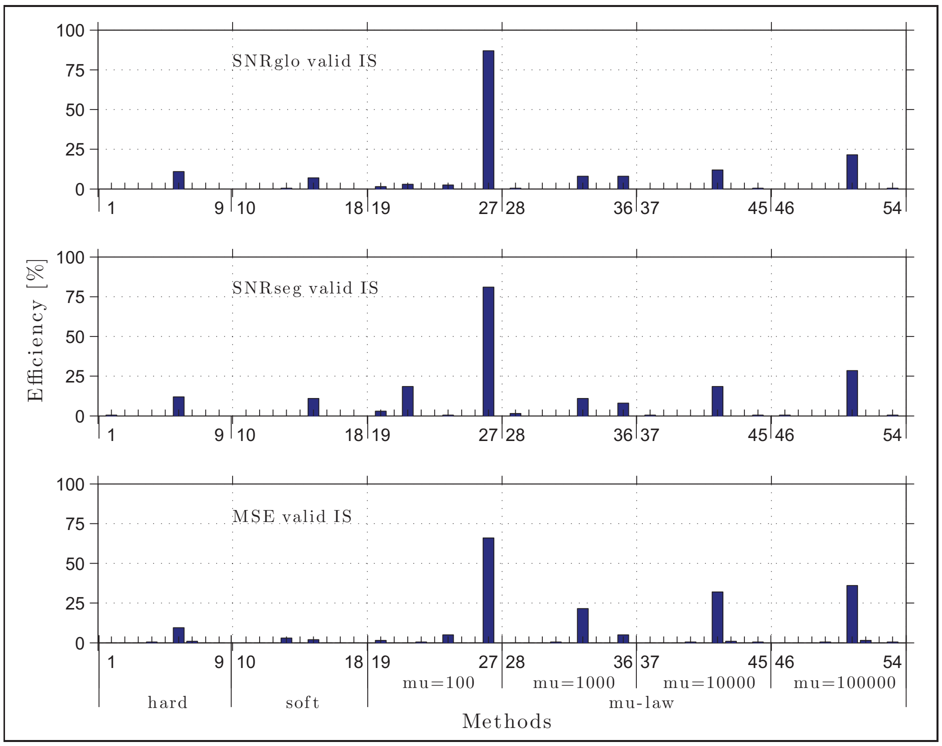

Figure 2.10 Third step experimental results: Selection of denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE.

2.5.4 Fourth step: Selection of the denoising techniques that simultaneously maximize performances for all the three criteria $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE

One technique seems to outperform all others in terms of efficiency as much for gain in $\mathrm{SNR_{glo}}$, gain in $\mathrm{SNR_{seg}}$ and MSE. This is technique No. 27 (defined in Table 2.2), which corresponds to the μ-law thresholding rule when $\mu = 100$, the hybrid $\mathrm{SURE_{soft}}$ threshold and the estimate of the standard deviation of the noise on each analysis level.

For the experimental signals considered, among the denoising methods considered and according to the selection criteria considered, the method that yields the best performance results is therefore the denoising using the Daubechies wavelet of order 4 with a 10 analysis levels and the use of the μ-law thresholding rule when $\mu = 100$ with the hybrid $\mathrm{SURE}_{\mathrm{soft}}$ threshold and the estimate of the standard deviation of the noise on each analysis level.

The actual performances in term of gain in$\mathrm{SNR}_{\mathrm{glo}}$, gain in $\mathrm{SNR}_{\mathrm{seg}}$, MSE and IS obtained by the application of this method are presented on the table 2.3. The first two noises are those extracted from the NOISEX database, the eight followed are those recorded at the NORANDA CCR copper refinery. The results are presented for the noised signal of -20, 0 and 20 dB SNR. Althougth the results obtained for all the noises tested are not exactly the same since some noises are more difficult to treat than others, some general observations can be made. The behavior of these quantities (gain in $\mathrm{SNR}_{\mathrm{glo}}$, gain in $\mathrm{SNR}_{\mathrm{seg}}$, MSE and IS) vary according to the $\mathrm{SNR}_{\mathrm{in}}$ of the input signal more or less in the same way for all the noises tested For all noises, it is easier to have high gain in $\mathrm{SNR}_{\mathrm{glo}}$ and $\mathrm{SNR}_{\mathrm{seg}}$ at low $\mathrm{SNR}_{\mathrm{in}}$ rather than at high $\mathrm{SNR}_{\mathrm{in}}$: at high $\mathrm{SNR}_{\mathrm{in}}$ the noise level is very low, so it is very difficult to suppress noise without suppressing speech, this is also why the MSE is higher at high $\mathrm{SNR}_{\mathrm{in}}$ than at low $\mathrm{SNR}_{\mathrm{in}}$. On the other hand the speech intelligibility (IS) is more easily deteriorated at low $\mathrm{SNR}_{\mathrm{in}}$ rather than at high $\mathrm{SNR}_{\mathrm{in}}$.

Table 2.3 Performance results for the best method for each noise

Criteria		IS			MSE			Gain in $\mathrm{SNR}_{\mathrm{glo}}$			Gain in $\mathrm{SNR}_{\mathrm{seg}}$		
$\mathrm{SNR}_{\mathrm{in}}$ [dB]		-20	0	20	-20	0	20	-20	0	20	-20	0	20
Noise	**1**	5.0	1.7	0.1	0.2	0.5	0.8	8.2	2.9	1.3	10.3	3.8	1.8
	2	2.4	0.4	0.0	0.1	0.2	0.4	10.4	6.4	4.4	12.1	6.3	4.4
	3	1.9	0.2	0.0	0.0	0.1	0.2	16.9	11.3	6.7	17.9	12.5	8.4
	4	2.4	0.6	0.0	0.0	0.2	0.3	15.8	8.1	5.0	16.8	8.4	5.9
	5	2.2	0.4	0.0	0.0	0.1	0.8	15.8	9.1	1.3	16.6	10.0	3.9
	6	2.2	0.3	0.0	0.1	0.2	0.3	9.2	7.1	5.0	12.6	7.7	5.3
	7	3.6	1.5	0.1	0.0	0.4	0.7	16.0	4.4	1.4	16.7	6.0	1.8
	8	3.0	1.4	0.1	0.0	0.4	0.9	13.4	3.9	0.6	14.8	5.1	1.1
	9	1.9	0.3	0.0	0.1	0.2	0.5	12.8	6.2	3.2	14.4	6.9	3.8
	10	3.3	1.5	0.1	0.1	0.4	0.8	12.6	4.2	1.1	14.6	5.1	1.6

2.6 Discussion

In this section, firstly, the general methodology developped to obtain our results is discussed. Secondly, the results are discussed. Thirdly, the independence or interdependence of parameters is examined. Finally, we will attempt to answer the following question: Which theoretical Gaussian noise (white or pink) best simulates industrial noises in a speech denoising treatment ?

2.6.1 Methods selection methodology: A positioning among similar studies

We shall begin by positioning our study among similar studies already presented in the literature and which we mentioned in the introduction. Antoniadis et al. (Antoniadis et al., 2001) tested 34 wavelet denoising methods on a dozen of theoretical signals altered by white Gaussian noise. Their study is mostly qualitative and does not provide a systematic comparison of the results according to the theoretical signals considered. Fodor and Kamath (Fodor and Chandrika, 2003) conducted the same type of study by denoising images altered by a white Gaussian noise. They consider 36 denoising by wavelet methods and 20 spatial filter techniques. For each set of methods, they presented a global table of results, which they later used to interpret the influence of various parameters. The study by Ayat et al. (Ayat et al., 2006) focuses on a large number of wavelet denoising methods tested on speech signals altered by white Gaussian noise. They took into consideration 22 wavelet types, 4 numbers of analysis levels, 3 threshold expressions and 5 thresholding rules. Their study was based on the variation of one parameter at a time.

Instead of the traditional white Gaussian noise (Antoniadis et al., 2001; Ayat et al., 2006; Fodor and Chandrika, 2003) used in the literature, we use industrial noises recorded in factories to be as realistic as possible. In addition, the range of SNRs considered is broad [-20 20] dB and subdivided into 1 dB increments, whereas the range of SNRs studied in the literature is seldom lower than -10 dB and often contains less than 10 different values (Ayat et al., 2006; Fodor and Chandrika, 2003). As shown in Table 2.1, 1 296 denoising methods are considered for this study; hence a global study analysis as the one conducted by Fodor (Fodor and Chandrika,

2003) on 36 techniques would be difficult and lengthy to perform. A parameter by parameter study, such as the one conducted by Ayat et al. (Ayat et al., 2006) might have been possible, however, we chose to proceed otherwise to avoid affecting our results with the mathematical dependence of certain parameters on other ones (see Section 2.2). We were able to study the influence of the analysis wavelet type and of the number of analysis levels independently because they are independent from the other parameters (see Sections 2.5.1.1 and 2.5.1.2). However the thresholding rule, the threshold expression and the noise estimate expression are not independent parameters, so they were studied simultaneously (see Sections 2.5.2, 2.5.3 and 2.5.4) using our selection algorithm (see Section 2.4.5). Among the four selection criteria considered, $\mathrm{SNR}_{\mathrm{glo}}$, SNE_{seg}, MSE and IS (see Section 2.3.3), we have chosen to favour speech intelligibility, which is quantified by the IS distortion measure because a method that would give very good gain in terms of SNR yet for which the denoised signal would be incomprehensible would not be useful for our purposes.

2.6.2 Analysis of experimental results

First we will discuss hereafter the choice of the analysis wavelet type. Then the number of analysis levels chosen will be discussed. Finally we will examine the choices made in terms of the parameters used in the denoising techniques (thresholding rule, threshold expression and noise estimate expression) in the third section.

2.6.2.1 Choice of the analysis wavelet type

Our study enables us to bring to evidence (see Section 2.5.1.1) that the Daubechies wavelets of order 4 provides good denoising results with a reasonably short processing time, while the Daubechies wavelets of order 1 gives slightly poorer results. We will try hereafter to give an explanation to these observations based upon the correspondence between the analysis level and the central frequency of the wavelet analysis bands. These central frequencies are shown in Fig. 2.11 for each tested wavelet type. On a logarithmic scale, excluding Daubechies wavelet of order 1, all wavelets have almost equal central frequencies, which differ by more or less 1 percent. The central frequencies for Daubechies wavelet of order 1 are shifted by more than

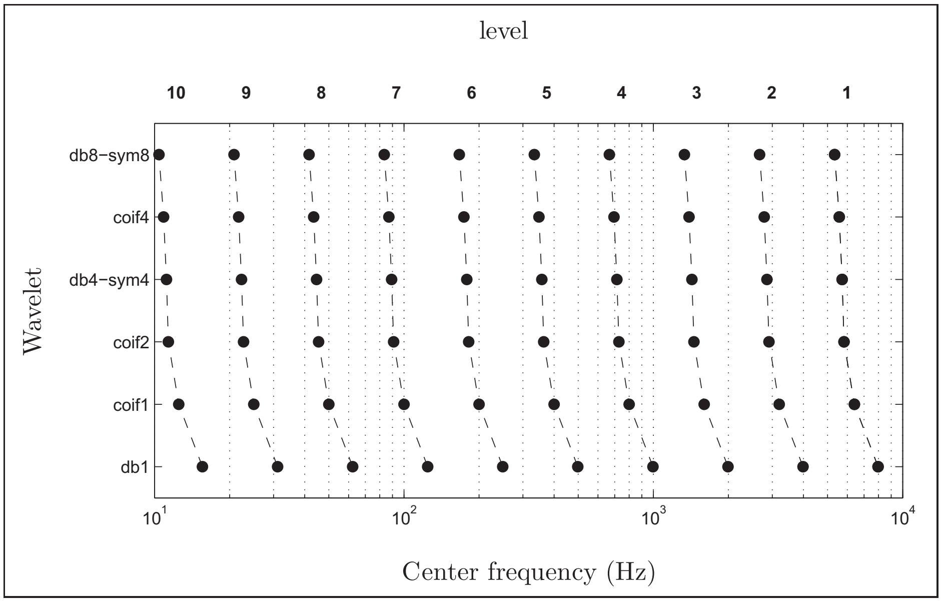

Figure 2.11 Correspondence between the level of decomposition and the central frequency for each tested wavelet type.

3.5%. Thus, the Daubechies wavelet of order 1 is somewhat less accurate in their representation of lower frequencies than the other wavelets at a given analysis level. Since industrial noises often contain a significant amount of low frequencies, lower denoising performances could be expected from its use than from the use of other wavelets (see Fig. 2.7). We can deduce from the study of Fig. 2.11 that, aside from the Daubechies wavelet of order 1, all the other considered wavelets (db4, db8, sym4, sym8, coif1, coif2 and coif4) are almost equivalent with respect to their respective central frequencies positions. The choice of the Daubechies wavelets of order 4 is justified by its wide use in the literature, its ease of implementation relatively to the Coiflets and Symlets wavelets and its low filter length requirements.

2.6.2.2 Choice of the number of analysis levels

The analysis of our results (see Section 2.5.1.2) shows that denoising by wavelet conducted on 10 analysis levels gives the best results. When denoising by wavelets, the detail coefficients are thresholded while the approximation of lower frequencies remains the same. Industrial noises often contain a significant range of low frequencies. To ensure "complete" denoising of the speech signal, the entire range of audible frequencies (20 Hz - 20 kHz) must be processed. Among the three analysis levels tested (6, 8 and 10), only analysis level 10 provided the necessary accuracy in the lower frequencies (see Fig. 2.11).

2.6.2.3 Selection of the denoising techniques

Steps 2 to 4 of our general methodology (see Fig. 2.2) enabled us to show that, for all the simulations performed, the μ-law thresholding rule with $\mu = 100$, hybrid SURE threshold and noise estimate on each analysis level, provides the best results (see Sections 2.5.2, 2.5.3 and 2.5.4). According to Fig. 2.1, this rule is, among all the rules presented in Table 2.1, the one that decreases the least below the threshold, the values of the wavelet coefficients. This results in a lower amount of deterioration of the speech signal (Nordström et al., 1999; Sheikhzadeh and Abutalebi, 2001). Also, the hybrid SURE threshold is intermediate between the universal threshold and the SURE threshold; in fact, by design, it increases low SURE threshold values obtained for low SNRs (Donoho and Johnstone, 1995). This choice is completely justified here since the technique(s) are chosen to ensure performance over the entire range of the considered SNRs. Finally, estimation of the standard deviation of the noise on each analysis level is meant for coloured noises, while estimation solely on the first analysis level is meant for white Gaussian noise. This points toward the possibility that industrial noises are more similar to coloured noises than white Gaussian noise, this will be further examined in section 2.6.4.

2.6.3 Observations on independent and interdependent parameters

The necessity to design an algorithm to study the performance of only the 54 denoising techniques and not of all the 1 296 denoising methods considered has been justify mathematically (see Section 2.4): some parameters and independent but not others. Here the results are examined in order to observe the dependence and interdependence of parameters. First the dependence of the analysis wavelet type and the number of analysis level from each other and from other parameters is examined. Second the interdependence of the denoising techniques parameters (the thresholding rule, the threshold expression and the noise estimate expression) is examined.

2.6.3.1 Independent parameters: Analysis wavelet type and number of analysis levels

The analysis wavelet type and the number of analysis levels are mathematically independent parameters both from each other and from the other denoising parameters (thresholding rule, threshold expression and noise estimate expression) (see Section 2.2).

Instead of using the selection algorithm only on the 54 denoising techniques as in the general methodology (see Section 2.4), the selection algorithm is applied here on all the 1 296 denoising methods. First the denoising methods that preserve intelligibility according to the IS distortion measure are determined, second the denoising methods that preserve intelligibility and that separately maximize the global SNR, maximize the segmental SNR and minimize the MSE are determined. This means that the second and third steps of the general methodology (see Section 2.4) are performed without taking into account the results of the first step. Two histograms are obtained, they are similar to the ones in Figs. 2.9 and 2.10 (see Sections 2.5.2 and 2.5.3), but they contain 1 296 bars (methods) instead of 54 bars (methods), hence they will not be presented in this paper but the main observations to which they lead will be reported hereafter.

On the one hand, the Daubechies wavelet of order 1 gives slightly poorer results compared to other wavelets, which all give similar results regardless of the selection criteria considered. In the first step of our general methodology, the same conclusion has been obtained, as it can be seen on Fig. 2.7 (see Section 2.5.1.1). On the other hand, better results are obtained with ten analysis levels than with eight analysis levels, and the use of only six analysis levels yields poor results. The same conclusion has also been obtained in the first step of our general methodology, as it can be deduced from Fig. 2.8 (see Section 2.5.1.2).

These observations confirm that the analysis wavelet type and the number of analysis levels are mutually independent. They are both also independent from the other denoising parameters.

2.6.3.2 Interdependent denoising technique parameters: Thresholding rule, threshold expression and noise estimate expression

The three denoising technique parameters thresholding rule, threshold expression and noise estimate expression, are mathematically interdependent: the thresholding rule is based on the threshold, which is in its turn based on the noise estimate (see Section 2.2). The results obtained at the second and third step of our general methodology, as shown in Figs. 2.9 and 2.10 (see Sections 2.5.2 and 2.5.3), will be examined hereafter in more details.

Firstly, let us consider the results of step 2 of our general methodology presented on Fig. 2.9 (see Section 2.5.2). Among the 54 denoising techniques considered (see Tables 2.1 and 2.2 in Section 2.3), we have decided to retain only those that preserve intelligibility of the speech signal (see Section 2.4.2). Thus, the selected techniques are those for which the results from step 2 using the selection algorithm are nonzero: intelligibility is preserved for at least 1 speech-noise pair. Let us examine for example the techniques that were not selected (for which the efficiency is zero according to the IS criterion). Depending on the thresholding rule considered, the denoising techniques that are not selected are not the same. For example, it is only for hard and soft thresholding rules, that the technique using the hybrid SURE threshold with a noise estimate on each analysis level is not selected (see on Fig. 2.9: methods No. 9 and 18 are not selected whereas methods No. 27, 36, 45 and 54 are selected). However, for all thresholding rules, the technique that uses the hybrid SURE threshold with noise estimate on the first analysis level is never selected (see results for methods No. 8, 17, 26, 35, 44, 53 on Fig. 2.9) and the technique that uses the hybrid SURE threshold with noise estimate by 1 is always selected (see results for methods No. 8, 16, 25, 34, 43, 52 on Fig. 2.9). So with a given threshold expression (hybrid SURE threshold in this example), the choice of the noise estimate expression and the choice of the thresholding rule are interdependent. In the same way, we can observe on Fig. 2.9 that for a given noise estimate expression (on each analysis level for example), the choice of the threshold expression and the choice of the thresholding rule are interdependent (see on Fig. 2.9: methods No. 3, 9, 12, 18, 30, 38 and 48 are not selected whereas methods No. 6, 15, 21, 24, 27, 33, 35, 42, 45, 51 and 54 are selected). Hence, the observations made on Fig. 2.9 indicate as expected that the three denoising parameters are interdependent.

Secondly, let us look at the results from step 3 of our general methodology on Fig. 2.10 (see Section 2.5.3). With the three selection criteria, the denoising technique, with μ-law thresholding rule when $\mu = 100$, the hybrid SURE threshold and the noise estimate on each analysis level (see results for method No. 27 on Fig. 2.10), is always at maximum efficiency regardless of the selection criteria considered. However, for the other thresholding rules, it must be noted that for the techniques also using the hybrid SURE threshold and estimate of the standard deviation of the noise on each analysis level (see results for methods No. 9, 18, 36, 45 and 54 on Fig. 2.10), efficiency is much lower, even zero. For these other thresholding rules, it is on the contrary the SURE threshold with noise estimate on each analysis level (see results for methods No. 6, 15, 33, 42, 51 on Fig. 2.10) that is the most efficient one. So the choice of the threshold expression depends on the choice of the thresholding rule.

In conclusion, these observations confirm the mathematical interdependence of the three denoising parameters.

2.6.4 Industrial noises versus white and pink Gaussian noises

All the methods of speech denoising by wavelets tested on industrial noises were also tested on white and pink Gaussian noises for comparative purposes. The results presented below are those obtained using a Daubechies wavelet of order 4 decomposed into 10 analysis levels. Fig. 2.12 shows the results obtained for white Gaussian noise. Fig. 2.13 shows the results obtained for pink Gaussian noise. These results were obtained in the same manner as for the industrial noises, using the same general methodology. The figures shown are to be compared with Fig. 2.10 (see Section 2.5.3), which shows the results obtained for the industrial noises.

A simple comparison of Figs. 2.12 and 2.13, which correspond to the results obtained for white and pink Gaussian noises, indicates that denoising speech altered by these two types of theoretical noises cannot be conducted in the same manner. White Gaussian noise appears to

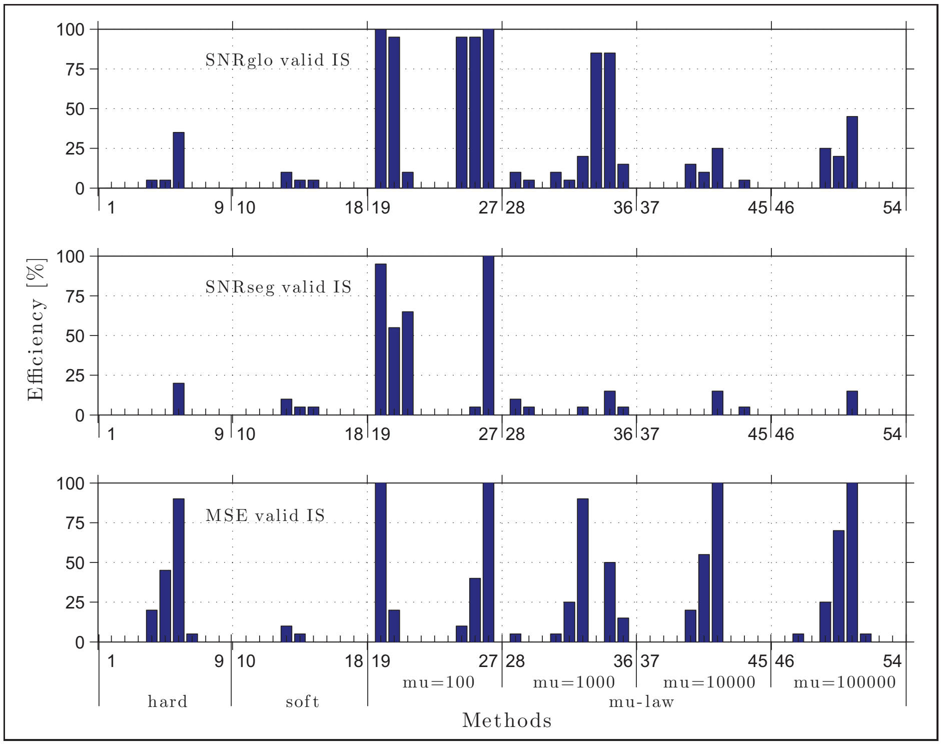

Figure 2.12 Third step experimental results: Selection of denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE for white Gaussian noise.

be easier to remove than pink Gaussian noise, since 2 techniques (No. 19 and 27) provide an efficiency superior to 75% for the three criteria $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE simultaneously. Furthermore, if we look at the criteria separately, there are respectively 7 (No. 19, 20, 25, 26, 27, 34 and 35), 2 (No. 19 and 27) and 6 (No. 6, 19, 27, 33, 42 and 51) number of techniques which give an efficiency superior to 75% for $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE. However, for the pink Gaussian noise, with the same selection criteria, only 1 method (No. 27) provided an efficiency superior to 75% for the three criteria used simultaneously or separately.

Comparing Figs. 2.12 and 2.13 to Fig. 2.10, which are all obtained with the same criteria (minimum quality, absolute deviation and relative deviation parameters, presented in section 2.4.5), many differences can be observed between Figs. 2.12 and Fig. 2.10, whereas Figs. 2.13 and 2.10 have a great similarity: (i) the same general pattern is obtained, (ii) most of the methods which have a significant efficiency are the same, (iii) the method which gives the best efficiency for all the three criteria is the same. Consequently, in the framework of the

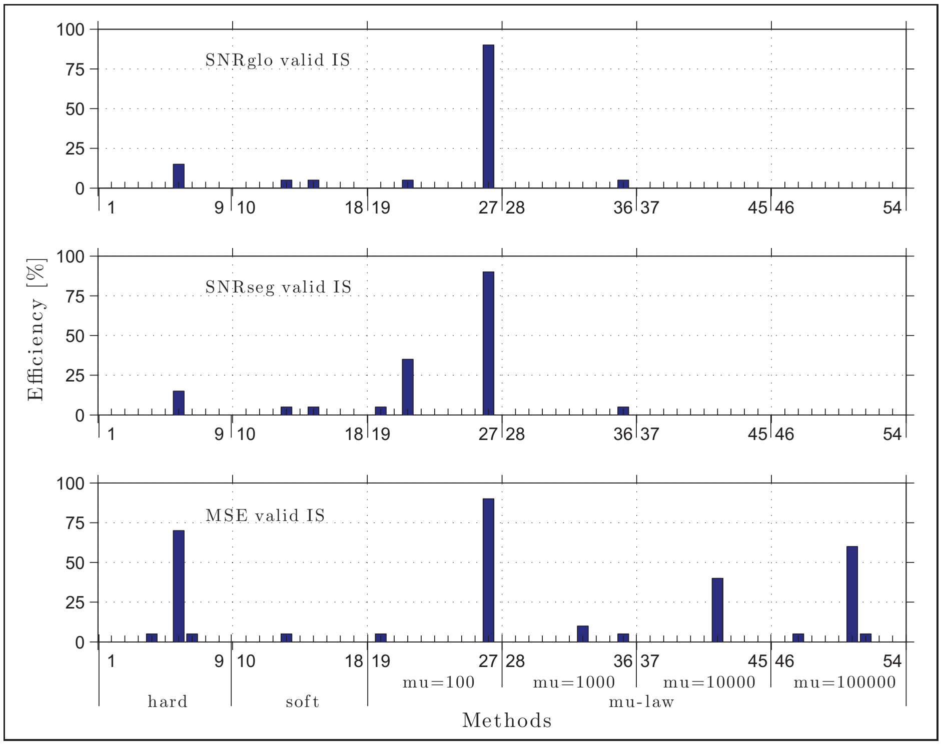

Figure 2.13 Third step experimental results: Selection of denoising techniques that separately optimize each criterion $\mathrm{SNR}_{\mathrm{glo}}$, $\mathrm{SNR}_{\mathrm{seg}}$ and MSE for pink Gaussian noise.

simulations performed, for a denoising speech treatment, industrial noises are more similar to the theoretical pink Gaussian noise than to the white one.

2.7 Conclusions and recommendations

In a research framework that aims at enabling industry worker wearing hearing protectors to communicate with others, this study has addressed the problem of denoising speech signals in an industrial noise environment so that the intelligible speech signal could be electrically transmitted inside the hearing protector. For this purpose, speech denoising methods by classical wavelet thresholding applied to an industrial noise environment have been studied. A large number (1 296) of denoising methods by classical wavelet thresholding, defined by 8 analysis wavelet types, 3 values of the number of analysis levels, 6 thresholding rules, 3 threshold expressions and 3 noise estimate expressions, have been considered. These 1 296 denoising methods have been applied to 8 200 noised speech signals, obtained by adding 20 speech signals to 10 industrial noises according to 41 signal to noise ratios. So, $1\ 296 \times 8\ 200$ denoised speech signals have been obtained. Their performances results have been determined according to 4 selection criteria which are: the global and the segmental signal to noise ratios, the mean square error and the Itakura-Saito distortion measure. Hence a database of 42 508 800 values has been generated and analysed.

A parameter by parameter approach could have been used to analyze these results. However, because of the interdependence of the three denoising parameters (thresholding rule, threshold expression and noise estimate expression), we have chosen to use a global methodology. Due to the large amount of data to be analyzed, a selection algorithm had to be developed. To start our procedure, the adequate analysis wavelet type and number of analysis levels have been determined independently. Then the selection algorithm has been used to determine the most efficient technique(s) according to the four selection criteria.

The best results for all the considered denoising techniques, over the entire set of experimental signals tested and for the 4 selection criteria used ($\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$, MSE and IS) has been obtained by using the following parameters: Daubechies wavelet of order 4 on 10 analysis levels, μ-law thresholding rule ($\mu = 100$), hybrid SURE threshold and noise estimate on each analysis level.

In addition, this study has indicated that the stationary-like industrial noise used to perform it, behave with respect to the denoising techniques which we have used, in a way quite similar to that of the pink Gaussian noise. This suggest that just as white Gaussian noise is used as a standard for denoising in telecommunications, pink Gaussian noise should be considered in further studies to be used as a standard for denoising in industrial noise environments for noises that tend to be stationary.

In this study, some problems have been encountered in the determination of the SURE threshold for the hard and the μ-law thresholding. The estimator of the risk associated to the hard thresholding is a biased one and it was difficult to obtain an estimator of the risk associated to the μ-law thresholding. In order to circumvent this problem, we have chosen to use the SURE

threshold associated to the soft thresholding for the hard and the μ-law thresholding. Other avenues do exist and could be explored in future works. One of them consists in obtaining a SURE threshold specifically optimized for the μ-law thresholding. This could improve the performance results of speech denoising.

Acknowledgements

The authors wish to express their appreciation to SONOMAX HEARING HEALTHCARE INC. (Montréal, Canada) and to NSERC (Natural Sciences and Engineering Research Council of Canada) for their support in this project.

References

Antoniadis, A., Bigot, J., Sapatinas, T., 2001. Wavelet estimator in nonparametric regression: a comparative simulation study. Journal of Statistical Software 6 (6), 1–83.

Ayat, S., Manzuri-Shalmani, M., Dianat, R., 2006. An improved wavelet-based speech enhancement by using speech signal features. Computers and Electrical Engineering 32 (6), 411–425.

Berger, E., Royster, L., Royster, J., Driscoll, D., Layne, M., 2000. The Noise Manual, 5th Edition. American Industrial Hygiene Association (AIHA) Press.

Bui, T., Chen, G., 1998. Translation-invariant denoising using multiwavelets. Signal Processing, IEEE Transactions on 46 (12), 3414–3420.

Donoho, D. L., Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81 (3), 425–455.

Donoho, D. L., Johnstone, I. M., 1995. Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association 90 (432), 1200–1225.

Ephraim, Y., 1992. Statistical-model-based speech enhancement systems. Proceedings of the IEEE 80 (10), 1526–1555, tY - JOUR.

Fan, G., Xia, X.-G., 2001. Improved hidden markov models in the wavelet-domain. Signal Processing, IEEE Transactions on 49 (1), 115–120.

Fodor, I. K., Chandrika, K., 2003. Denoising through wavelet shrinkage: an empirical study. Journal of Electronic Imaging 12 (1), 151–160.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., Zue, V., 1993. Timit acoustic-phonetic continuous speech corpus, No. LDC93S1, Linguistic Data Consortium, Philadelphia.

Gustafsson, H., Nordholm, S., Claesson, I., 2001. Spectral subtraction using reduced delay convolution and adaptive averaging. Speech and Audio Processing, IEEE Transactions on 9 (8), 799–807, tY - JOUR.

Hansen, J. H., Pellom, B., 1998. An effective quality evaluation protocol for speech enhancement algorithms. In: ICSLP-98: Inter. Conf. on Spoken Language Processing. Vol. 7. Sydney, Australia, pp. 2819–2822.

Haykin, S., 2002. Adaptive filter theory, 4th Edition. Prentice-Hall information and system sciences series. Prentice-Hall, Upper Saddle River, N.J., États-Unis.

Johnstone, I. M., Silverman, B. W., 1997. Wavelet threshold estimators for data with correlated noise. Journal of the Royal Statistical Society Series B-Methodological 59 (2), 319–351.

Krim, H., Tucker, D., Mallat, S., Donoho, D. L., 1999. On denoising and best signal representation. Information Theory, IEEE Transactions on 45 (7), 2225–2238.

Lu, C.-T., Wang, H.-C., 2007. Speech enhancement using hybrid gain factor in critical-band-wavelet-packet transform. Digital Signal Processing 17 (1), 172–188.

Noisex-92, 1990, Institute for Perception-TNO, The Netherlands, Speech Research Unit, RSRE, United Kingdom.

Nordström, F., Holst, B., Lindoff, B., 1999. Time and frequency dependent noise reduction in speech signals. In: The International Conference on Signal Processing Applications and Technology. Orlando, Florida, USA.

Sheikhzadeh, H., Abutalebi, H. R., 2001. An improved wavelet-based speech enhancement system. In: Eurospeech 2001. Vol. 3. Aalborg, Denmark, Scandinavia, pp. 1855–1858.

Wang, S., Sekey, A., Gersho, A., 1992. An objective measure for predicting subjective quality of speech coders. Selected Areas in Communications, IEEE Journal on 10 (5), 819–829.

WHO, 1999. Guidelines for community noise. Tech. rep.

WHO, 2001. Occupational exposure to noise: evaluation, prevention and control. Tech. rep., Federal Institute for Occupational Safety and Health.

CHAPITRE 3

ARTICLE #3

"A WAVELET SPEECH THRESHOLDING RULE FOR DENOISING IN INDUSTRIAL ENVIRONMENTS"

Cécile Le Cocq, Christian Gargour, Frédéric Laville,
École de technologie supérieure, Université du Québec,
Montréal (Québec), Canada, H3C 1K3
This article has been submitted to IEEE Transaction on Signal Processing 29 September 2009.

Résumé

Le débruitage par ondelettes est appliqué dans cet article à la solution du problème de communication associé au port de protecteurs auditifs par les travailleurs en milieux bruyants. Une nouvelle règle de seuillage est proposée. Une expression mathématique compacte de l'estimateur du risque associé à cette règle a été obtenue. Il est démontré que cet estimateur n'est pas biaisé. Les performances obtenues par l'utilisation de cette règle de seuillage ont été testées et évaluées en utilisant différentes combinaisons additives de paroles et de bruits industriels sur une vaste plage de rapports signal à bruit et en utilisant différents critères. Les performances obtenues semblent être égales ou supérieures à celles données par la règle de seuillage basée sur la loi de μ.

Abstract

Wavelet based denoising is applied in this paper to the solution of the communication problem associated with the hearing protectors used by workers operating in noisy environments. A new thresholding rule designed for wavelet speech denoising in industrial environments is proposed. A closed form mathematical expression for the risk estimator associated with this rule is obtained and is shown to be unbiased. Its performances are tested and evaluated using various speech and industrial noise combinations over a wide range of signal to noise ratios

and using several criteria. The obtained performances appear to be superior or equal to those given by the μ-law thresholding rule.

3.1 Introduction

In order to avoid hearing damages, workers operating in noisy industrial environments have to wear hearing protectors which could make speech communication difficult. This situation leads them to remove temporarily their protectors whenever they have to communicate verbally with each other. By doing so, they expose themselves to noise levels which could harm their hearing. Speech wavelet denoising in industrial environments is considered in this paper as a tool which could be used to deal with this problem. Since the concept of industrial noise has not, to our knowledge, been yet formally defined in the literature, we have considered for our purpose, that any noise to which a worker is exposed in an industrial workplace is an industrial noise.

A large number of studies on speech denoising by wavelet thresholding for telecommunications purposes have been reported in the literature (Antoniadis et al., 2001; Fodor and Chandrika, 2003; Ayat et al., 2006; Tantibundhit et al., 2007; Yu and Chip-Hong, 2007). Applications of this technique to noisy industrial environments, are less abundant. In a previous paper on that topic (Le Cocq et al., tted), we have studied the performances of some classical wavelet based denoising methods in industrial environments. This paper has led to the conclusion that for the considered noisy industrial environments the μ-law thresholding rule used with the hybrid SURE (Stein Unbiased Risk Estimator) threshold with noise estimation performed on each analysis level, gives the best performances of speech denoising for the considered methods and criteria framework. In this previous work however, the SURE threshold of the soft thresholding rule has been used in the case of the μ-law thresholding rule. A mathematical expression for the corresponding risk estimator cannot be easily derived (Le Cocq et al., tted). In this paper, we propose a new thresholding rule which will be designated here as the CLC thresholding rule. It is designed such as an unbiased risk estimator mathematical expression can be obtained for it. This rule is studied and evaluated.

Consider a signal such as:

$$x = s + w \tag{3.1}$$

where s is the speech signal and w is the noise. The wavelet thresholding consist in applying a thresholding rule THR with the threshold T on the wavelet coefficients :

$$X = S + W \tag{3.2}$$

of the signal x in order to obtain the denoised speech $\tilde{S}$:

$$\tilde{S} = \mathrm{THR}\,(X, T) \tag{3.3}$$

The five major thresholding rules reported in the literature are the following : hard, soft (Donoho and Johnstone, 1994), hard-soft, super-soft (Nordström et al., 1999) and μ-law (Sheikhzadeh and Abutalebi, 2001). Table 3.1 presents some of the main advantages and inconveniences which results from the mathematical properties of these rules. Other thresholding rules have been reported in the literature, such as non-negative garrote (Breiman, 1995), firm (semisoft) and garrote (Gao, 1998), custom (Yoon and Vaidyanathan, 2004), super super soft and hard super soft (Nordström, 1998), step-garotte (Ayat et al., 2006), SCAD and Bayesian (Antoniadis et al., 2001), etc. They have advantages and inconveniences similar to the aforementioned ones. We will propose in this paper the CLC thresholding rule which does not have any of the inconveniences appearing in Table 3.1: It is continuous, it does not set to zero any of the wavelet coefficients and it keeps unchanged the values of the large ones. Also the risk estimator of this rule can be obtained in the form a mathematical expression and it is unbiased.

The paper is organized as follows. Section 3.2 presents the proposed thresholding rule and its associated risk estimator. The speech and noise signals, together with all the other wavelet denoising parameters and the performance evaluation criteria which have been utilized to test this thresholding rule are covered in section 3.3. Section 3.4 exposes the experimental results. Our conclusions are given in section 3.5. Appendixes 3.A and 3.B contain the mathematical derivations pertaining to this work.

Table 3.1 Advantages and inconveniences of some of the major thresholding rules reported in the literature

Thresholding rule	Mathematical properties	Advantages (+) & Inconveniences (-)
Hard	$\tilde{S} = 0$ if $\|X\| < T$	(-) Decreased intelligibility at these levels (Nordström et al., 1999; Sheikhzadeh and Abutalebi, 2001)
	$\tilde{S} = X$ if $\|X\| \geq T$	(+) Unaltered intelligibility at these levels
	Discontinuity at $\|X\| = T$	(-) Biased risk estimator: poorer denoising performances (Krim et al., 1999)
Soft	$\tilde{S} = 0$ if $\|X\| < T$	(-) Decreased intelligibility at these levels (Nordström et al., 1999; Sheikhzadeh and Abutalebi, 2001)
	$\tilde{S} < X$ if $\|X\| \geq T$	(-) Decreased intelligibility at these levels (Nordström et al., 1999)
	Continuity	(+) Unbiased risk estimator
Hard-soft	$\tilde{S} > 0$ if $0 < \|X\| < T$	(+) Unaltered intelligibility at these levels
	$\tilde{S} = X$ if $\|X\| \geq T$	(+) Unaltered intelligibility at these levels
	Discontinuity at $\|X\| = T$	(-) Biased risk estimator: poorer denoising performances (Krim et al., 1999)
Super-soft	$\tilde{S} > 0$ if $0 < \|X\| < T$	(+) Unaltered intelligibility at these levels
	$\tilde{S} < X$ if $\|X\| \geq T$	(-) Decreased intelligibility at these levels (Nordström et al., 1999)
	Continuity	(+) Unbiased risk estimator
μ-law	$\tilde{S} > 0$ if $0 < \|X\| < T$	(+) Unaltered intelligibility at these levels
	$\tilde{S} = X$ if $\|X\| \geq T$	(+) Unaltered intelligibility at these levels
	Continuity	(-) Quite difficult to obtain the risk estimator (Le Cocq et al., tted)

3.2 Proposed method

3.2.1 Thresholding rule

The proposed CLC thresholding rule does not modify the wavelet coefficients which are superior to the threshold T. Between 0 and T, the rule is divided into two linear segments. The slopes of these two segments are respectively α and $1/\beta$. For $|X| \geq T$, the signal remains unchanged. The values of the two parameters α and β are constrained between 0 and 1.

$$\mathrm{THR}_{CLC}(X,T) = \begin{cases} \alpha X & \text{if } |X| < T_0 \\ \frac{1}{\beta}X - \mathrm{sgn}(X)\frac{1-\beta}{\beta}T & \text{if } T_0 \leq |X| < T \\ X & \text{if } |X| \geq T \end{cases} \tag{3.4}$$

with $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ and $T_0 = \frac{1-\beta}{1-\alpha\beta}T$.

So the new thresholding rule is continuous. It does not eliminate any wavelet coefficient. Also the high valued coefficients remain unchanged.

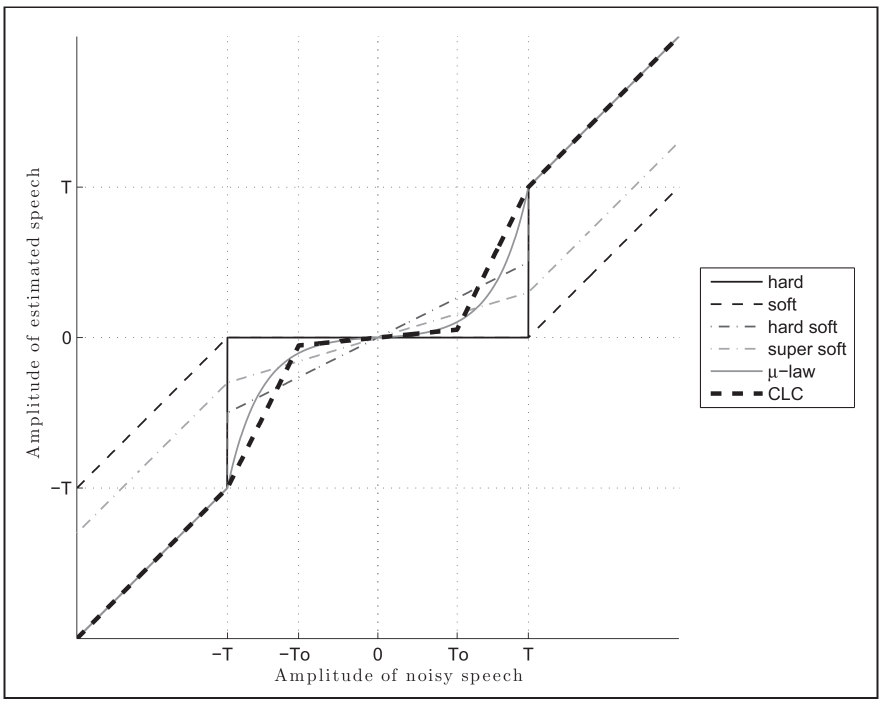

Figure 3.1 Thresholding rules.

Figure 3.1 illustrates the thresholding rules reported in the literature as well as the CLC rule.

3.2.2 Risk estimator

The risk $r_{\mathrm{THR}}(s,T)$ associated to a thresholding rule THR with a given threshold T is defined by:

$$r_{\mathrm{THR}}(s,T) = \mathrm{E}\left\{\left\|S - \tilde{S}\right\|^2\right\} \tag{3.5}$$

The risk estimator for the new thresholding rule can be expressed in a compact mathematical form as follows (the derivation is given in appendix 3.A):

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s,T) = (l-p)\left(\frac{1-\beta}{\beta}\right)^2 T^2 + \left[(N-l) + (l-p)\frac{2-\beta}{\beta} - p(1-2\alpha)\right]\sigma^2 - 2\left(\frac{1-\beta}{\beta}\right)^2 T \sum_{T_0 \leq |X| < T} |X[k]| + \left(\frac{1-\beta}{\beta}\right)^2 \sum_{T_0 \leq |X| < T} |X[k]|^2 + (1-\alpha)^2 \sum_{|X| < T_0} |X[k]|^2 \tag{3.6}$$

where N is the length of the speech signal, p is the number of samples $X[m]$ which verify $|X[m]| < T_0$ $(p = \#\{m : |X[m]| < T_0\})$, similarly $l = \#\{m : |X[m]| < T\}$.

It can be demonstrated (cf. appendix 3.B) that this estimator is unbiased, i.e.:

$$r_{\mathrm{THR}_{CLC}}(s,T) = \mathrm{E}\left\{\tilde{r}_{\mathrm{THR}_{CLC}}(s,T)\right\} \tag{3.7}$$

3.3 Framework for the evaluation of the proposed method

Table 3.2 presents all the considered cases in this study. The next three subsections present the different components of our testing procedure: signals, denoising parameters and performance criteria.

Table 3.2 Signals, denoising parameters and criteria

Signals	Speech:	- 10 sentences (male)
		- 10 sentences (female)
	Noise:	- 2 car factory noises from Noisex
		- 8 Noranda copper refinery noises
	SNR:	from -20 dB to 20 dB in 1 dB steps
Denoising parameters	- Wavelet Daubechies db4	
	- 10 analysis levels	
	- CLC thresholding rule	$\alpha = 0.1, 0.3, 0.5, 0.7$
		$\beta = 0.1, 0.3, 0.5, 0.7$
	- Universal, SURE and hybrid SURE thresholds	
	- Noise estimate on each analysis level	
Performance criteria	- Global signal to noise ratio ($\mathrm{SNR}_{\mathrm{glo}}$)	
	- Segmental signal to noise ratio ($\mathrm{SNR}_{\mathrm{seg}}$)	
	- Mean square error (MSE)	
	- Itakura-Saito distortion measure (IS)	

3.3.1 Signals

Twenty english speech signals extracted from the TIMIT database (Garofolo et al., 1993) are altered by ten industrial noises: two from the NOISEX database (NOISEX , 1990) and eight recorded at the NORANDA CCR copper refinery of Montréal (J. Voix, personal communication, 2000) cf. Table 3.3, according to 41 SNRs (signal to noise ratio) distributed between -20 dB and 20 dB as follows:

$$x = \frac{s}{\mathrm{std}(s)} 10^{\mathrm{SNR}/20} + \frac{w}{\mathrm{std}(w)} \tag{3.8}$$

where the standard deviation of the signal s given by:

$$\mathrm{std}(s) = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} \Big(s[n] - \tilde{s}[n] \Big)^2} \tag{3.9}$$

In this way a set of $20 \times 10 \times 41 = 8200$ noised speech signal are formed and considered.

Table 3.3 Noises

Index	Name	Description
1	NOISEX 21	Car Factory: car floor production, electrical welding
2	NOISEX 22	Car Factory: car production hall
3	NORANDA CCR 01	Copper refinery: Room packing selenium - Baghouse #1
4	NORANDA CCR 02	Copper refinery: Room packing selenium - Baghouse #1, #2, #3 and Vaccum cleaner - 82 dBA
5	NORANDA CCR 03	Copper refinery: Room packing selenium - Selenium micropulverisator + Baghouse ventilator - 79 dBA
6	NORANDA CCR 04	Copper refinery: Room packing selenium - Baghouse security valve - 71 dBA
7	NORANDA CCR 05	Copper refinery: Room packing selenium - Booster with rotary rammers (Hertxen 1812)
8	NORANDA CCR 06	Copper refinery: Powerhouse #2 - Transformers room - 100 dBA
9	NORANDA CCR 07	Copper refinery: Tankhouse - Propane burner, furnace #8 - 102 dBA
10	NORANDA CCR 08	Copper refinery: Hydraulic room - Dumm press - 108 dBA

3.3.2 Methods

The wavelet decomposition is performed on 10 levels with the Daubechies wavelet of order 4. The CLC thresholding rule is tested with four possibles values (0.1, 0.3, 0.5 and 0.7) of each of the two parameters α and β. The noise estimate is determined on each analysis level i by $\tilde{\sigma}_i = \mathrm{MAD}(X_i)/0.6745$ with $\mathrm{MAD}(X_i) = \mathrm{median}(|X_i|)$ (Johnstone and Silverman, 1997). The proposed thresholding rule is tested with three threshold expressions (Donoho and Johnstone, 1994, 1995; Johnstone and Silverman, 1997): the universal threshold $T_u(i) = \tilde{\sigma}_i\sqrt{2\log N}$, the SURE threshold $T_{\mathrm{SURE}}(i) = \tilde{\sigma}_i \, \mathrm{SURE}(X_i/\tilde{\sigma}_i)$ with $\mathrm{SURE}\,(X) = \arg\min_{0\leq T} \tilde{r}_{\mathrm{THR}}(s, T)$, and the hybrid SURE threshold $T_{\mathrm{SURE\ hybrid}}$ defined by:

$$T_{\mathrm{SURE\ hybrid}} = \begin{cases} T_{\mathrm{SURE}} & \text{if } \|X\|^2 - N\sigma^2 \leq \epsilon_N \\ T_u & \text{if } \|X\|^2 - N\sigma^2 > \epsilon_N \end{cases} \tag{3.10}$$

where $\epsilon_N = \sigma^2 N^{1/2} (\log N)^{3/2}$

Hence a set of $4 \times 4 \times 3 = 48$ methods are tested. They are enumerated and labeled in Table 3.4. This table is used for easy reference to these 48 methods in the remaining of the paper.

3.3.3 Criteria

Four criteria (Hansen and Pellom, 1998; Wang et al., 1992) have been chosen to quantify the performance of the proposed thresholding rule: the global signal to noise ratio $\mathrm{SNR}_{\mathrm{glo}} = 10\log_{10}\{\mathrm{var}(s)/\mathrm{var}(s-\bar{s})\}$, the segmental signal to noise ration $\mathrm{SNR}_{\mathrm{seg}} = 1/M\sum_{m=0}^{M-1} 10\log_{10}\{\mathrm{var}(s_m)/\mathrm{var}(s_m-\bar{s}_m)\}$, the mean square error $\mathrm{MSE} = 1/N\sum_{n=0}^{N-1}(s[n]-\bar{s}[n])^2$, and the Itakura-Saito distortion measure:

$$\mathrm{IS} = \frac{1}{M}\sum_{m=0}^{M-1}\left\{\left[\frac{G^2_{s_m}}{G^2_{\tilde{s}_m}}\right]\left[\frac{\vec{a}_{\tilde{s}_m}\mathbf{R}_{s_m}\vec{a}^T_{\tilde{s}_m}}{\vec{a}_{s_m}\mathbf{R}_{s_m}\vec{a}^T_{s_m}}\right] + \log\left(\frac{G^2_{\tilde{s}_m}}{G^2_{s_m}}\right) - 1\right\} \tag{3.11}$$

This last criterion can be used (Hansen and Pellom, 1998; Wang et al., 1992) to quantify the level of intelligibility of a speech signal.

Table 3.4 Denoising techniques

Index	**Thresholding parameters**	**Threshold**
1		universal
2	$\alpha = 0.1$ & $\beta = 0.1$	$SURE_{clc}$
3		hybrid $SURE_{clc}$
4		universal
5	$\alpha = 0.3$ & $\beta = 0.1$	$SURE_{clc}$
6		hybrid $SURE_{clc}$
7		universal
8	$\alpha = 0.5$ & $\beta = 0.1$	$SURE_{clc}$
9		hybrid $SURE_{clc}$
10		universal
11	$\alpha = 0.7$ & $\beta = 0.1$	$SURE_{clc}$
12		hybrid $SURE_{clc}$
13		universal
14	$\alpha = 0.1$ & $\beta = 0.3$	$SURE_{clc}$
15		hybrid $SURE_{clc}$
16		universal
17	$\alpha = 0.3$ & $\beta = 0.3$	$SURE_{clc}$
18		hybrid $SURE_{clc}$
19		universal
20	$\alpha = 0.5$ & $\beta = 0.3$	$SURE_{clc}$
21		hybrid $SURE_{clc}$
22		universal
23	$\alpha = 0.7$ & $\beta = 0.3$	$SURE_{clc}$
24		hybrid $SURE_{clc}$
25		universal
26	$\alpha = 0.1$ & $\beta = 0.5$	$SURE_{clc}$
27		hybrid $SURE_{clc}$
28		universal
29	$\alpha = 0.3$ & $\beta = 0.5$	$SURE_{clc}$
30		hybrid $SURE_{clc}$
31		universal
32	$\alpha = 0.5$ & $\beta = 0.5$	$SURE_{clc}$
33		hybrid $SURE_{clc}$
34		universal
35	$\alpha = 0.7$ & $\beta = 0.5$	$SURE_{clc}$
36		hybrid $SURE_{clc}$
37		universal
38	$\alpha = 0.1$ & $\beta = 0.7$	$SURE_{clc}$
39		hybrid $SURE_{clc}$
40		universal
41	$\alpha = 0.3$ & $\beta = 0.7$	$SURE_{clc}$
42		hybrid $SURE_{clc}$
43		universal
44	$\alpha = 0.5$ & $\beta = 0.7$	$SURE_{clc}$
45		hybrid $SURE_{clc}$
46		universal
47	$\alpha = 0.7$ & $\beta = 0.7$	$SURE_{clc}$
48		hybrid $SURE_{clc}$

3.4 Experimental results

The selection algorithm given in Le Cocq et al. (tted) has been used here. It avoids the biasing of the results which could occur as a consequence of the interdependence of the following parameters: the thresholding rule, the threshold expression and the noise estimate expression.

The first step of our method consists in determining the efficiency for the IS distortion measure of each of the denoising techniques listed in Table 3.4. This efficiency is the percentage of speech-noise pairs for which a minimum level of intelligibility heuristically determined (Le Cocq et al., tted) and quantified by the IS distortion measure is obtained. The efficiency of all the considered techniques are represented by the histogram appearing in Figure 3.2. In this figure, the numbers appearing on the horizontal axis refer to the index column of Table 3.4.

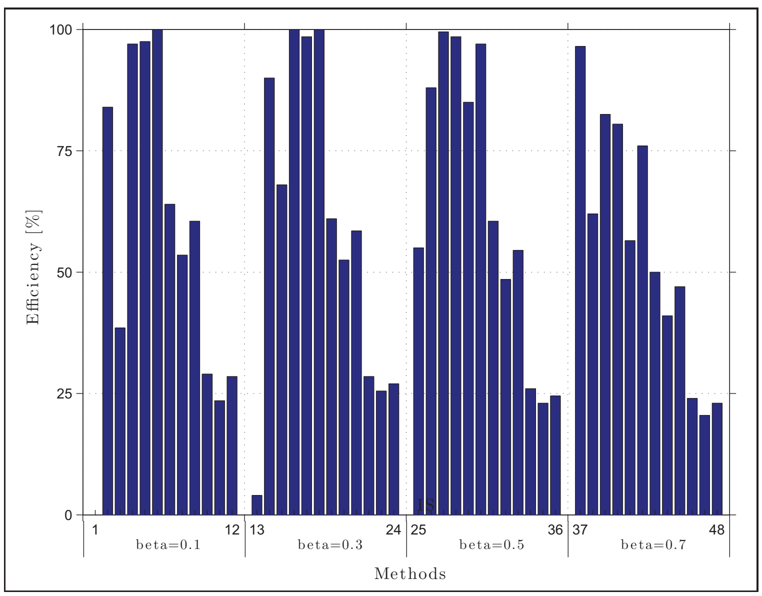

Figure 3.2 Selection of denoising techniques that preserve intelligibility.

In the second step of our method, for each speech-noise pair, only the techniques for which at least the minimum level of intelligibility is reached, are retained. For these techniques only, the selection algorithm is applied to the results corresponding to the three other criteria namely the $\mathrm{SNR}_{\mathrm{glo}}$, the $\mathrm{SNR}_{\mathrm{seg}}$ and the MSE. The corresponding efficiencies are determined. These efficiencies are respectively depicted by the three histograms of Figure 3.3. Here also the numbers appearing on the horizontal axis refer to the index column of Table 3.4. It appears

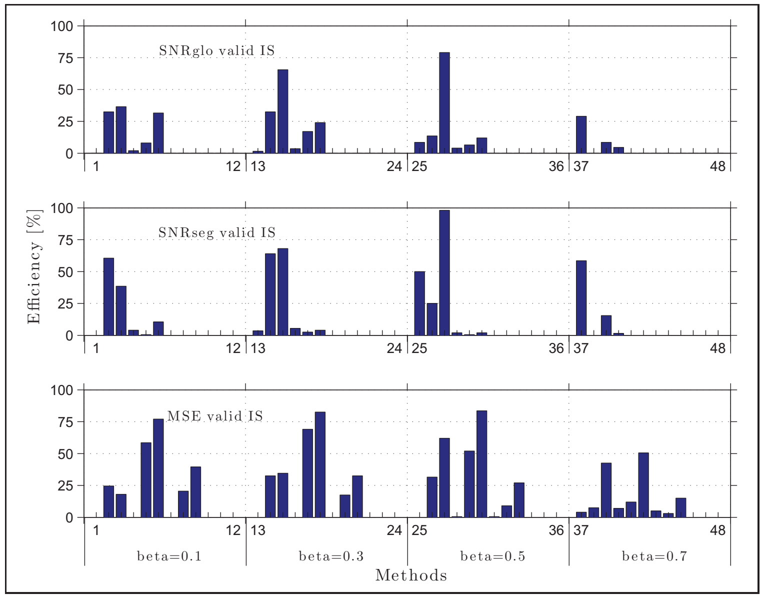

Figure 3.3 Selection of denoising techniques that separately optimize each criterion $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE.

from the graph, that the technique which gives the best compromise between efficiencies for $\mathrm{SNR_{glo}}$, $\mathrm{SNR_{seg}}$ and MSE is the one which corresponds to index 27. This technique makes use of the hybrid SURE threshold expression optimized for the proposed thresholding rule with $\alpha = 0.1$ and $\beta = 0.5$.

Table 3.5 Performance results for the best method for each noise

Criteria		IS			MSE			Gain in $\mathrm{SNR_{glo}}$			Gain in $\mathrm{SNR_{seg}}$		
$\mathrm{SNR_{in}}$ [dB]		-20	0	20	-20	0	20	-20	0	20	-20	0	20
Noise	**1**	3.9	1.5	0.2	0.2	0.4	1.3	8.1	3.9	-1.0	11.3	6.6	1.1
	2	2.3	0.5	0.1	0.1	0.3	0.8	8.9	5.9	0.9	11.3	7.4	1.8
	3	1.6	0.3	0.1	0.0	0.1	0.7	16.2	10.8	2.1	17.5	13.0	4.2
	4	2.4	0.9	0.2	0.0	0.2	0.8	15.1	8.3	1.2	16.4	10.6	3.3
	5	1.9	0.4	0.2	0.0	0.1	1.2	15.1	8.7	-0.6	16.5	11.2	2.3
	6	2.2	0.4	0.1	0.2	0.3	0.8	7.6	5.9	1.0	11.8	7.9	1.5
	7	3.8	1.6	0.3	0.0	0.3	1.3	15.4	5.0	-1.0	16.3	9.0	1.8
	8	3.2	1.6	0.3	0.1	0.4	1.4	11.7	4.4	-1.5	14.0	8.2	1.4
	9	1.8	0.5	0.1	0.1	0.3	1.0	10.7	6.0	0.2	13.7	8.8	2.1
	10	3.4	1.4	0.2	0.1	0.3	1.4	10.8	4.6	-1.2	13.9	7.9	1.1

The performance of all the speech-noise pairs have been averaged for each of the considered noises numbered from one to ten according to Table 3.3. The corresponding results are given in Table 3.5 for each of the four criteria (IS, MSE, $\mathrm{SNR_{glo}}$ and $\mathrm{SNR_{seg}}$) and for three signal to noise ratios (-20 dB, 0 dB and 20 dB).

Figure 3.4 shows the gain in segmental signal to noise ratio of the μ-law thresholding rule (Le Cocq et al., tted) and of the CLC one. In each of the two cases, the denoising is performed using the corresponding hybrid-SURE threshold expression. It appears from the figure that for signal to noise ratios ranging from -10 dB to 15 dB, the CLC thresholding rule gives better results than the μ-law one. The corresponding comparison curves for the three other criteria are not represented here. They show however no significant differences between the two considered methods.

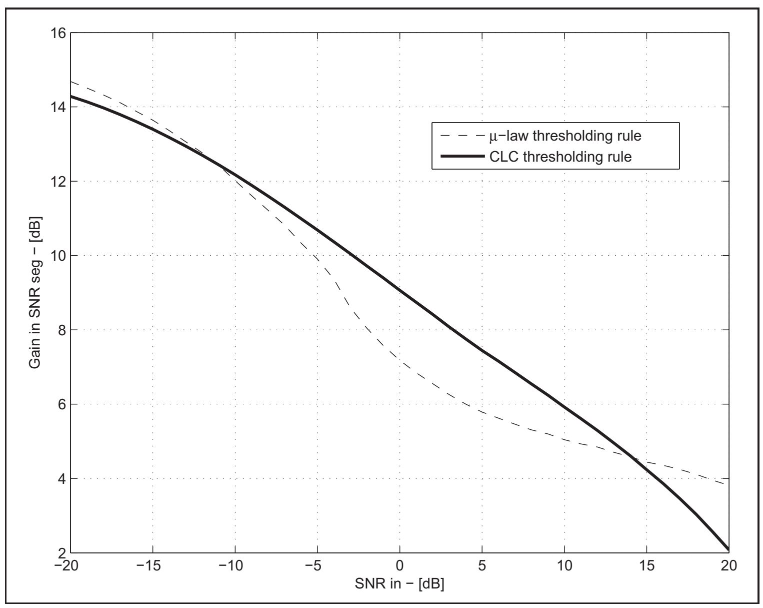

Figure 3.4 Denoising performances, in term of the gain in segmental SNR, for the μ-law and the CLC thresholding rule, according to different signal to noise ratios.

3.5 Conclusions

In this paper, speech denoising by means of wavelet thresholding in industrial environments has been considered. While industrial noise are not formally defined in the literature, we have used an assortment of such noises which we consider representative of the noises which can be found in industries. Since it has been shown in a previous study that the use of the μ-law thresholding rule appears to be adequate for our purpose but that its risk estimator cannot be easily mathematically expressed, we have developed a new one by approximating its curve by two linear segments of respective slopes α and β. A compact mathematical expression for the risk estimator of the proposed thresholding rule has been obtained. It has been shown that this estimator is unbiased. Different values of the slopes α and β have been tested on various speech signals altered by several industrial noises with a wide range of signal to noise ratios. This have been done for the universal, SURE and hybrid-SURE threshold expressions. The criteria which have been used for this purpose are the Itakura-Saito distortion measure, the segmental and global signal to noise ratio and the mean square error. Among the values of the slopes α and β and the threshold expressions yielding to a minimum acceptable level of intelligibility according to the Itakura-Saito distortion measure, we have selected the combinations of slopes and threshold expression which achieve a good compromise of the three other performance criteria, namely the global and segmental signal to noise ratio and the mean square error. The experimental results obtained by our approach appear in general to be as good as, or better than, the ones given by the μ-law thresholding rule.

3.A Derivation of the risk estimator for the proposed thresholding rule

The definition of the risk given by equation 3.5 and the relation between S, W, and X given by equation 3.2 are both applied to the three forms of the proposed thresholding rule presented in equation 3.4.

- In the case where: $|x| < T_0$

$$r_{\mathrm{THR}_{CLC}}(s, T) = \sum_{m=1}^{N} \mathrm{E}\left\{\left|S[m] - \alpha X[m]\right|^2\right\}$$

hence,

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s, T) = \sum_{m=1}^{N} \left[(1-\alpha)^2\, \mathrm{E}\left\{|X[m]|^2\right\} - (1-2\alpha)\sigma^2\right] \tag{3.12}$$

- In the case where: $T_0 \leq |x| < T$

$$r_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N} \mathrm{E}\left\{\left|S[m] - \left[\frac{1}{\beta}X[m] - \frac{1-\beta}{\beta}T\,\mathrm{sgn}\left(X[m]\right)\right]\right|^2\right\}$$
$$= \sum_{m=1}^{N} \mathrm{E}\left\{\left|\left(1-\frac{1}{\beta}\right)S[m] - \frac{1}{\beta}W[m] + \frac{1-\beta}{\beta}T\,\mathrm{sgn}\left(X[m]\right)\right|^2\right\}$$

Assuming that $S[m]$ and $X[m]$ are not correlated and that $\mathrm{E}\{|W[m]|\} = 0$, the previous expression can be approximated as:

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N}\left[\left(\frac{1-\beta}{\beta}\right)^2 \mathrm{E}\{|X[m]|^2\} - 2\left(\frac{1-\beta}{\beta}\right)^2 T\,\mathrm{E}\{|X[m]|\} + \left(\frac{1-\beta}{\beta}\right)^2 T^2 + \left(\frac{2-\beta}{\beta}\right)\sigma^2\right] \quad (3.13)$$

- In the case where: $|x| \geq T$

$$r_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N} \mathrm{E}\left\{\left|S[m] - X[m]\right|^2\right\}$$

with N the length of the signal. Hence,

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N}\sigma^2 \quad (3.14)$$

The three cases considered previously (equations 3.14, 3.13 and 3.12) can be expressed in a compact form as follows:

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s,T) = \sum_{|X[m]|\geq T}\sigma^2 + \sum_{|X|<T_0}\left[(1-\alpha)^2|X[k]|^2 - (1-2\alpha)\sigma^2\right]$$
$$+ \sum_{T_0\leq|X|<T}\left[\left(\frac{1-\beta}{\beta}\right)^2|X[k]|^2 - 2\left(\frac{1-\beta}{\beta}\right)^2 T|X[k]| + \left(\frac{1-\beta}{\beta}\right)^2T^2 + \left(\frac{2-\beta}{\beta}\right)\sigma^2\right] \quad (3.15)$$

Using N the length of the signal, $p = \#\{m : |X[m]| < T_0\}$ and $l = \#\{m : |X[m]| < T\}$ we obtain the final result (equation 3.6):

$$\tilde{r}_{\mathrm{THR}_{CLC}}(s,T) = (l-p)\left(\frac{1-\beta}{\beta}\right)^2T^2 + \left[(N-l) + (l-p)\frac{2-\beta}{\beta} - p(1-2\alpha)\right]\sigma^2$$
$$-2\left(\frac{1-\beta}{\beta}\right)^2 T\sum_{T_0\leq|X|<T}|X[k]| + \left(\frac{1-\beta}{\beta}\right)^2\sum_{T_0\leq|X|<T}|X[k]|^2 + (1-\alpha)^2\sum_{|X|<T_0}|X[k]|^2$$

3.B Derivation of the bias of the risk estimator for the proposed thresholding rule

To show that the risk estimator of the proposed thresholding rule is unbiased, an expression for the risk, in which no approximation is used, will be obtained hereafter. It will be shown that this expression is identical to the one obtained in equation 3.15 for the risk estimator.

Starting from the definition of the risk given by equation 3.5 expressed as:

$$r_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N} \mathrm{E}\left\{|S[m] - \mathrm{THR}_{CLC}(X[m],T)|^2\right\}$$

and using equation 3.4 expressed as:

$$\mathrm{THR}_{CLC}(X[m],T) = X[m] - g_{CLC}(X[m])$$

where

$$g_{CLC}(X) = \Big[(1-\alpha)X\Big]\mathbf{1}_{|X|<T_0} + \left[\left(1-\frac{1}{\beta}\right)X + \frac{1-\beta}{\beta}T\,\mathrm{sgn}(X)\right]\mathbf{1}_{T_0\le|X|<T}$$

we obtain the expression of the risk estimator for the proposed thresholding rule in the form:

$$r_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N}\left[\mathrm{E}\left\{\Big(N[m]\Big)^2\right\} + \mathrm{E}\left\{\Big(g_{CLC}(X[m])\Big)^2\right\} - 2\,\mathrm{E}\left\{N[m]\,g_{CLC}(X[m])\right\}\right] \tag{3.16}$$

Each of the two last terms of this expression is developed separately as follows:

- In the second term $g_{CLC}(X[m])$ is replaced by its expression given previously. Also the mathematical expectation operator acts separately on each of the terms of the sum on which it was applied

$$\begin{aligned}\mathrm{E}\left\{\Big(g_{CLC}(X[m])\Big)^2\right\} &= (1-\alpha)^2\,\mathrm{E}\left\{X[m]^2\mathbf{1}_{|X[m]|<T_0}\right\} \\ &\quad + \left(1-\frac{1}{\beta}\right)^2\Bigg[\mathrm{E}\left\{X[m]^2\mathbf{1}_{T_0\le|X[m]|<T}\right\} \\ &\quad + T^2\,\mathrm{E}\left\{\mathbf{1}_{T_0\le|X[m]|<T}\right\} - 2T\,\mathrm{E}\left\{|X[m]|\mathbf{1}_{T_0\le|X[m]|<T}\right\}\Bigg]\end{aligned} \tag{3.17}$$

- In the third term, using the definition of the mathematical expectation we obtain:

$$\mathrm{E}\left\{N[m]\,\mathrm{g}_{CLC}(X[m])\right\} = \int_{-\infty}^{+\infty} N[m]\,\mathrm{g}_{CLC}\left(N[m]+S[m]\right)\phi(N[m])dN[m]$$

where:

$$\phi(u) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\left(u^2/2\sigma^2\right)}$$

The integration by part of the above expression gives

$$\begin{aligned}\mathrm{E}\left\{N[m]\,\mathrm{g}_{CLC}(X[m])\right\} = \sigma^2 \int_{-\infty}^{+\infty}\left(1-\frac{1}{\beta}\right)\mathbf{1}_{T_0\leq|X[m]|<T}\phi(N[m])dN[m] \\ +\,\sigma^2\int_{-\infty}^{+\infty}(1-\alpha)\mathbf{1}_{|X[m]|<T_0}\phi(N[m])dN[m]\end{aligned} \quad (3.18)$$

Again, according to the definition of the mathematical expectation:

$$\mathrm{E}\left\{N[m]\,\mathrm{g}_{CLC}(X[m])\right\} = \left(1-\frac{1}{\beta}\right)\sigma^2\ \mathrm{E}\left\{\mathbf{1}_{T_0\leq|X[m]|<T}\right\} + (1-\alpha)\sigma^2\,\mathrm{E}\left\{\mathbf{1}_{|X[m]|<T_0}\right\} \quad (3.19)$$

The expressions of equations 3.17 and 3.19 are replaced in equation 3.16. This gives:

$$\begin{aligned} r_{\mathrm{THR}_{CLC}}(s,T) = \sum_{m=1}^{N}\Bigg[\ \sigma^2\,\mathrm{E}\left\{\mathbf{1}_{|X[m]|\geq T}\right\} + \mathrm{E}\left\{\left[(1-\alpha)^2X[m]^2-(1-2\alpha)\sigma^2\right]\mathbf{1}_{|X[m]|<T_0}\right\} \\ +\,\mathrm{E}\left\{\left[\left(1-\frac{1}{\beta}\right)^2\left(X[m]^2+T^2-2T|X[m]|\right)+\frac{2-\beta}{\beta}\sigma^2\right]\mathbf{1}_{T_0\leq|X[m]|<T}\right\}\Bigg]\end{aligned} \quad (3.20)$$

Comparing equation 3.15 and equation 3.20, it can be noticed that the mathematical expectation $\mathrm{E}\left\{\tilde{r}_{\mathrm{THR}_{CLC}}(s,T)\right\}$ of the risk estimator of the proposed thresholding rule (cf. equation 3.15) is equal to $r_{\mathrm{THR}_{CLC}}(s,T)$ (cf. equation 3.20) which is the exact expression for the risk of the proposed thresholding rule.

Hence (equation 3.7):

$$r_{\mathrm{THR}_{CLC}}(s,T) = \mathrm{E}\left\{\tilde{r}_{\mathrm{THR}_{CLC}}(s,T)\right\}$$

and it can be concluded that the risk estimator for the proposed thresholding rule is unbiased.

Acknowledgments

The authors wish to express their appreciation to SONOMAX HEARING HEALTHCARE INC. (Montréal, Canada) and to NSERC (Natural Sciences and Engineering Research Council of Canada) for their support in this project. They also thank professor Jean-Marc Lina for his useful advices regarding the mathematical demonstrations.

References

Anestis Antoniadis, Jérémie Bigot, and Theofanis Sapatinas. Wavelet estimator in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6(6):1–83, 2001.

Saeed Ayat, M.T. Manzuri-Shalmani, and Roohollah Dianat. An improved wavelet-based speech enhancement by using speech signal features. *Computers and Electrical Engineering*, 32(6):411–425, 2006.

Leo Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, 1995.

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1225, 1995.

Imola K. Fodor and Kamath Chandrika. Denoising through wavelet shrinkage: an empirical study. *Journal of Electronic Imaging*, 12(1):151–160, 2003.

Hong-Ye Gao. Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4):469–488, 1998.

John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, Nancy L. Dahlgren, and Victor Zue. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Linguistic Data Consortium, Philadelphia, 1993.

John H.L. Hansen and Bryan Pellom. An effective quality evaluation protocol for speech enhancement algorithms. In *ICSLP-98: Inter. Conf. on Spoken Language Processing*, volume 7, pages 2819–2822, Sydney, Australia, 1998.

I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society Series B-Methodological*, 59(2):319–351, 1997.

Hamid Krim, Dewey Tucker, Stéphane Mallat, and David L. Donoho. On denoising and best signal representation. *Information Theory, IEEE Transactions on*, 45(7):2225–2238, 1999.

Cécile Le Cocq, Christian Gargour, and Frédéric Laville. Wavelet speech enhancement for industrial noise environments. *Speech Communication*, submitted 2008.

NOISEX-92. Institute for Perception-TNO, The Netherlands, Speech Research Unit, RSRE, United Kingdom, 1990.

F. Nordström. *Time and Frequency Dependent Noise Reduction in Speech Signals.* Master, Lund Institute of Technology, 1998.

F. Nordström, B. Holst, and B. Lindoff. Time and frequency dependent noise reduction in speech signals. In *The International Conference on Signal Processing Applications and Technology*, Orlando, Florida, USA, 1999.

Hamid Sheikhzadeh and Hamid Reza Abutalebi. An improved wavelet-based speech enhancement system. In *Eurospeech 2001*, volume 3, pages 1855–1858, Aalborg, Denmark, Scandinavia, 2001.

C. Tantibundhit, J.R. Boston, C.C. Li, J.D. Durrant, S. Shaiman, K. Kovacyk, and A. El-Jaroudi. New signal decomposition method based speech enhancement. *Signal Processing*, 87(11):2607–2628, 2007.

Shihua Wang, Andrew Sekey, and Allen Gersho. An objective measure for predicting subjective quality of speech coders. *Selected Areas in Communications, IEEE Journal on*, 10(5):819–829, 1992.

Byung-Jun Yoon and P.P. Vaidyanathan. Wavelet-based denoising by customized thresholding. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 2, pages 925–928, Philadelphia, Pennsylvanie, 2004.

Shao Yu and Chang Chip-Hong. A generalized time?frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(4):877–889, 2007.

CONCLUSION

Dans le cadre de cette thèse, le problème de la communication dans le bruit, du point de vue à la fois des locuteurs et des auditeurs, a été étudié.

D'une part pour les locuteurs, le port de protecteurs auditifs de type bouchons d'oreilles entraine une modification de la perception de sa propre voix. Cette modification est en partie dû à l'effet d'occlusion qui modifie la perception interne de notre propre voix. De plus, le locuteur a tendance à corriger sa voix pour palier à ces modifications. Ces corrections entrainent une diminution de l'intelligibilité de ses paroles pour l'auditeur. Dans l'article #1, une étude de l'effet d'occlusion dû à des protecteurs auditifs a été effectuée afin de mieux comprendre l'influence de l'effet d'occlusion sur la perception de notre propre voix.

D'autre part pour les auditeurs, la communication en milieu industriel bruité est difficile en raison du fort niveau sonore. Dans les articles #2 et #3, le débruitage de la parole en milieu industriel bruité a été étudié. Ainsi le signal de parole débruité pourrait être réémis sous les protecteurs auditifs des auditeurs afin que ceux-ci puissent parfaitement percevoir, reconnaître et comprendre les paroles du locuteur.

Chacune de ces deux parties est détaillées ci-après, avant de faire une synthèse des résultats obtenus dans le cadre de cette thèse.

Perception interne de sa propre voix : Article #1

La figure 4.1 présente un schéma récapitulatif de la démarche suivie, des résultats obtenues et des recommandations pour l'article #1 intitulé "Subjective characterization of earplugs' occlusion effect using an external acoustical excitation of the mouth cavity".

En se basant sur les connaissances déjà présentes dans la littérature, un nouveau schéma des chemins de perception interne de la voix a été proposé. Il permet de représenter explicitement les différents chemins de transmission de la voix à l'oreille interne. Une distinction a été faite entre la transmission de la voix par le corps (VBC - voice body conduction) et la transmission

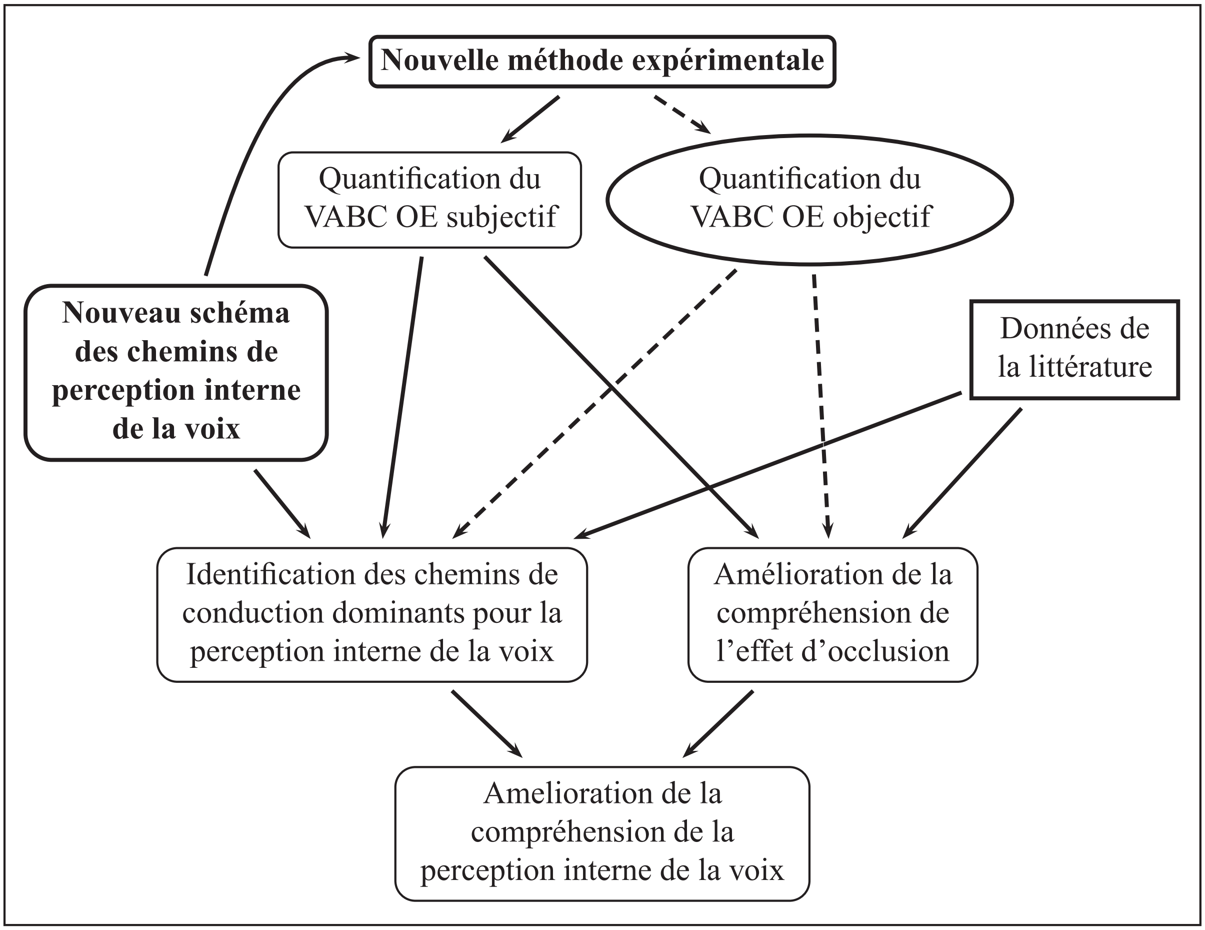

Figure 4.1 Schéma récapitulatif pour l'article #1

de la voix par l'air et le corps (VABC - voice air and body conduction). Le premier chemin (VBC) a pour origine les vibrations structurelles des cordes vocales, tandis que le deuxième chemin (VABC) a pour origine les vibrations de l'air dans le conduit buccal. Chacun de ces deux chemins se différencie en deux sous-chemins : le premier direct à l'oreille interne, le deuxième indirect qui passe par le conduit auditif externe avant de rejoindre l'oreille interne.

Dans la littérature, le chemin VBC a été simulé au moyen d'un vibrateur osseux et a permis des quantifications objectives et subjectives de l'effet d'occlusion correspondant. Des quantifications objectives de l'effet d'occlusion dû à la voix (somme des chemins VBC et VABC) sont aussi disponibles dans la littérature. Afin de compléter ces données de la littérature, une nouvelle méthode expérimentale pour quantifier l'effet d'occlusion dû au chemin VABC seul a été proposé dans cette thèse. Une source sonore (haut-parleur) a été placé de telle manière qu’elle excite la cavité buccale sans exciter l’oreille externe ou le reste du corps. Une quantification subjective de l’effet d’occlusion dû au chemin VABC a été obtenue. Les mesures de la quantification objective de l’effet d’occlusion dû au chemin VABC devront être repris en dehors de cette thèse en raison des problèmes techniques rencontrés.

L’analyse des ces différentes quantifications de l’effet d’occlusion a permis de tirer un certain nombre de conclusions sur l’effet d’occlusion et sur l’identification des chemins de transmission dominants pour la perception interne de la voix. En basses fréquences l’effet d’occlusion est toujours positif tandis qu’en hautes fréquences (dans les cas où une conclusion pouvait être tirée) il est négatif ou nul. En oreilles occluses en basses fréquences, c’est toujours le chemin indirect qui domine le chemin direct, tandis qu’en hautes fréquences le chemin direct domine le chemin indirect pour la transmission VBC. En oreille ouverte et en hautes fréquences, pour la transmission VBC c’est le chemin direct qui domine, alors que pour la transmission VABC c’est le chemin indirect qui domine.

L’étude réalisée dans le cadre de l’article #1 a donc permis l’identification des chemins de conduction dominants pour la perception interne de la voix, ainsi que l’amélioration de la compréhension de l’effet d’occlusion. Ces deux avancées scientifiques ont permis d’obtenir un amélioration de la compréhension de la perception interne de la voix.

Afin de compléter la recherche réalisée dans le cadre de l’article #1, il s’agirait tout d’abord de quantifier l’effet d’occlusion objectif dû au chemin de transmission VABC. Il serait également intéressant d’envisager de quantifier subjectivement l’effet d’occlusion dû à la voix. Il faudrait alors concevoir une méthode expérimentale qui n’utilise par la technique du seuil de perception. Des techniques de niveau équivalent ou de sonie équivalente pourraient être envisagées.

La recherche fondamentale effectuée dans l’article #1 permet de donner des pistes à d’éventuelles recherches appliquées aux prothèses auditives et aux protecteurs auditifs intra-auriculaires afin de compenser l’effet d’occlusion qui altère la perception de la propre voix du locuteur, et par conséquent l’intelligibilité de ses paroles pour l’auditeur.

Rehaussement de la parole en milieu industriel bruité : Articles #2 et #3

La figure 4.2 présente un schéma récapitulatif de la démarche suivie, des résultats obtenues et des recommandations pour l'article #2 intitulé “Wavelet speech enhancement for industrial noise environments”.

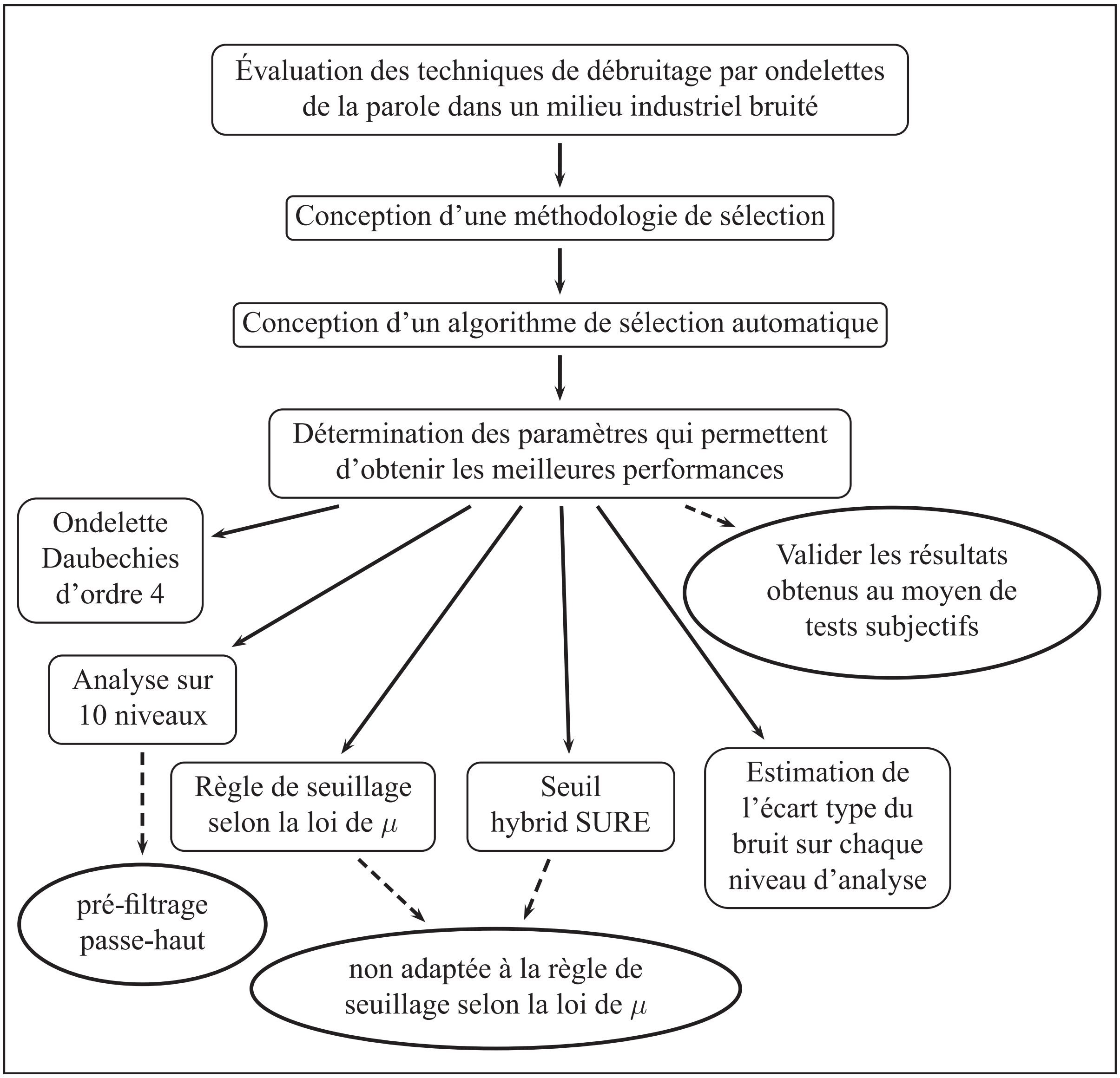

Figure 4.2 Schéma récapitulatif pour l'article #2

Une étude générale des performances du débruitage par ondelettes pour un signal de parole altéré par un bruit industriel a été réalisé dans l'article #2. Un ensemble de 8 200 signaux de parole bruités a été considéré (20 signaux de parole, 10 bruits industriels et 41 rapports signal à bruit). Ces signaux ont été débruités au moyen de 1296 méthodes de débruitages de la parole (8 ondelettes d'analyse, 3 nombres de niveaux d'analyse, 6 règles de seuillage, 3 calculs du seuil et 3 estimations de l'écart type du bruit). Quatre critères de sélection ont permis d'évaluer les performances du débruitage. Vu l'importance des données à traiter, une méthodologie de sélection et un algorithme de sélection automatique ont été conçus.

L'influence des différents paramètres du débruitage par ondelettes sur les performances obtenues a ainsi pu être mis en évidence et parmi les méthodes considérées, celle qui permet d'obtenir les meilleurs performances de débruitage pour l'ensemble des rapports signal à bruit a été identifiée :

– Le choix de l'ondelette n'a plus beaucoup d'influence sur les performances obtenues, dès qu'il s'agit d'une ondelette continue, le choix s'est donc porté sur l'ondelette Daubechies d'ordre 4.
– Les bruits industriels ayant de fortes composantes en basses fréquences, un nombre élevé (10) de niveaux d'analyse pour la transformée en ondelettes a été nécessaire pour optimiser les performances de débruitage.
– Les bruits industriels n'étant en général pas blanc, mais ayant au contraire un contenu spectral plutôt "coloré", réaliser l'estimation de l'écart type du bruit sur chaque niveau d'analyse a permis d'obtenir de meilleures performances.
– La règle de seuillage selon la loi de μ avec $\mu = 100$ a permis d'obtenir de meilleures performances car par opposition aux règles de seuillage mou ou dur, elle ne force aucun coefficients à zéro et ainsi dégrade moins l'intelligibilité de la parole.
– Le calcul du seuil par hybrid SURE a permis d'obtenir le meilleur compromis pour obtenir de bonnes performances de débruitage sur l'ensemble de la plage de rapport signal à bruit considérée.

Une méthode de débruitage de la parole par ondelettes adaptée au milieu industriel bruité a ainsi été déterminée de manière objective. Il serait intéressant d'effectuer des tests subjectifs afin de valider les résultats obtenus au moyen de l'algorithme de sélection. Par ailleurs, les bruits industriels ayant de fortes composantes basses fréquences, il serait intéressant de réaliser un pré-filtrage passe-haut afin de réduire la profondeur d'analyse de la transformée en ondelettes. Il est à noter également que le calcul du seuil hybrid SURE utilisé n'est pas adapté à la règle de seuillage selon la loi de μ, mais au seuillage mou.

Afin de remédier à cet inconvénient, dans le cadre de l'article #3, une nouvelle loi de seuillage a été proposée. La figure 4.3 présente un schéma récapitulatif de la démarche suivie, des résultats obtenues et des recommandations pour l'article #3 intitulé "A wavelet speech thresholding rule for denoising in industrial environments".

La nouvelle loi de seuilage vérifie les propriétés suivantes :

- Aucun coefficient n'est forcé à zéro.
- Les coefficients supérieurs au seuil restent inchangés.
- La loi de seuillage est continue en tout point.
- Il existe une forme mathématique compacte de l'estimateur du risque associé à cette loi.
- L'estimateur du risque associé à cette loi est non biaisé.

Une évaluation des performances a été effectuée pour différents paramètres du débruitage par seuillage d'ondelettes. Les paramètres qui permettent d'obtenir les meilleures performances de débruitage ont été mis en évidence :

- Comme dans l'article #2, l'expression du seuil hybrid SURE a permis d'obtenir les meilleures performances.
- Le paramètre $\alpha = 0.1$ de la loi de seuillage, qui correspond à la première pente après zéro, permet d'obtenir les meilleures performances. Il s'agit de la valeur du paramètre qui minimise le plus les faibles coefficients.
- Le paramètre $\beta = 0.5$ de la loi de seuillage, qui correspond à la deuxième pente après zéro, maximise les performances de débruitage.

L'étude des performances de cette nouvelle méthode de débruitage par seuillage d'ondelette a montré une amélioration des performances pour des rapports signal à bruit du signal bruité compris entre -10 dB et 15 dB par rapport à celles obtenues lors du débruitage par seuillage

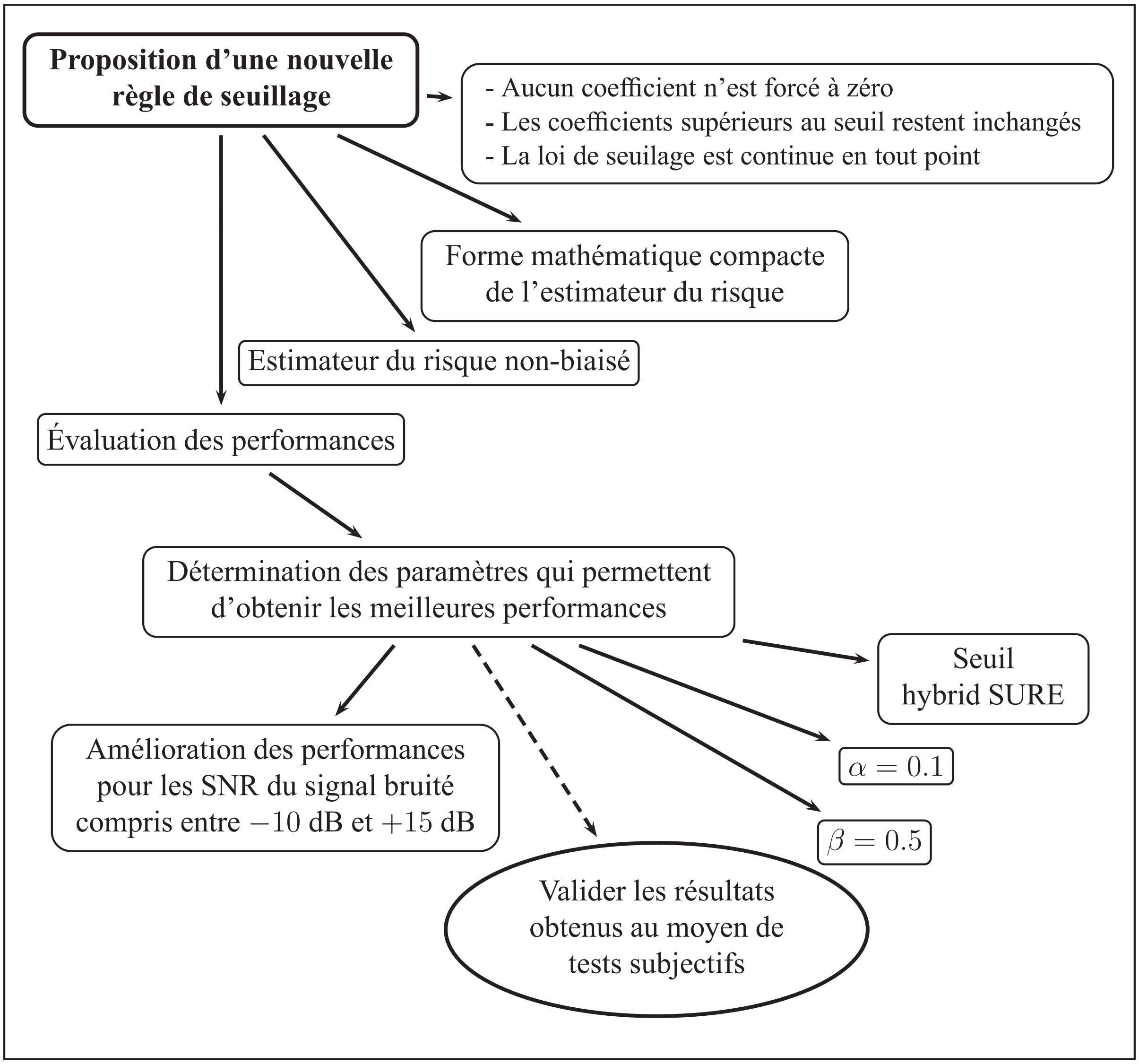

Figure 4.3 Schéma récapitulatif pour l'article #3

selon la loi de μ. De même que pour l'étude réalisé dans l'article #2, il serait pertinent de valider le résultats objectifs obtenus ici au moyen de tests subjectifs.

La méthode de débruitage par ondelettes ici proposée est une première solution pour les industriels qui voudraient inclure un débruitage de la parole dans leurs protecteurs auditifs. Les ouvriers auraient ainsi moins tendance à retirer leurs protecteurs pour parler avec leurs collègues.

Des améliorations sont envisageables sur cette méthode. Tout d'abord nous avons utilisée ici la transformée en ondelette classique. Afin de mieux simuler le comportement fréquentielle de la cochlée, il serait pertinent de remplacer cette transformée en ondelettes par une transformée en paquets d'ondelettes qui suivrait l'échelle de Bark ou de Mel.

Synthèse

La figure 4.4 présente un schéma récapitulatif de la démarche suivie, des résultats obtenues et des recommandations pour la thèse.

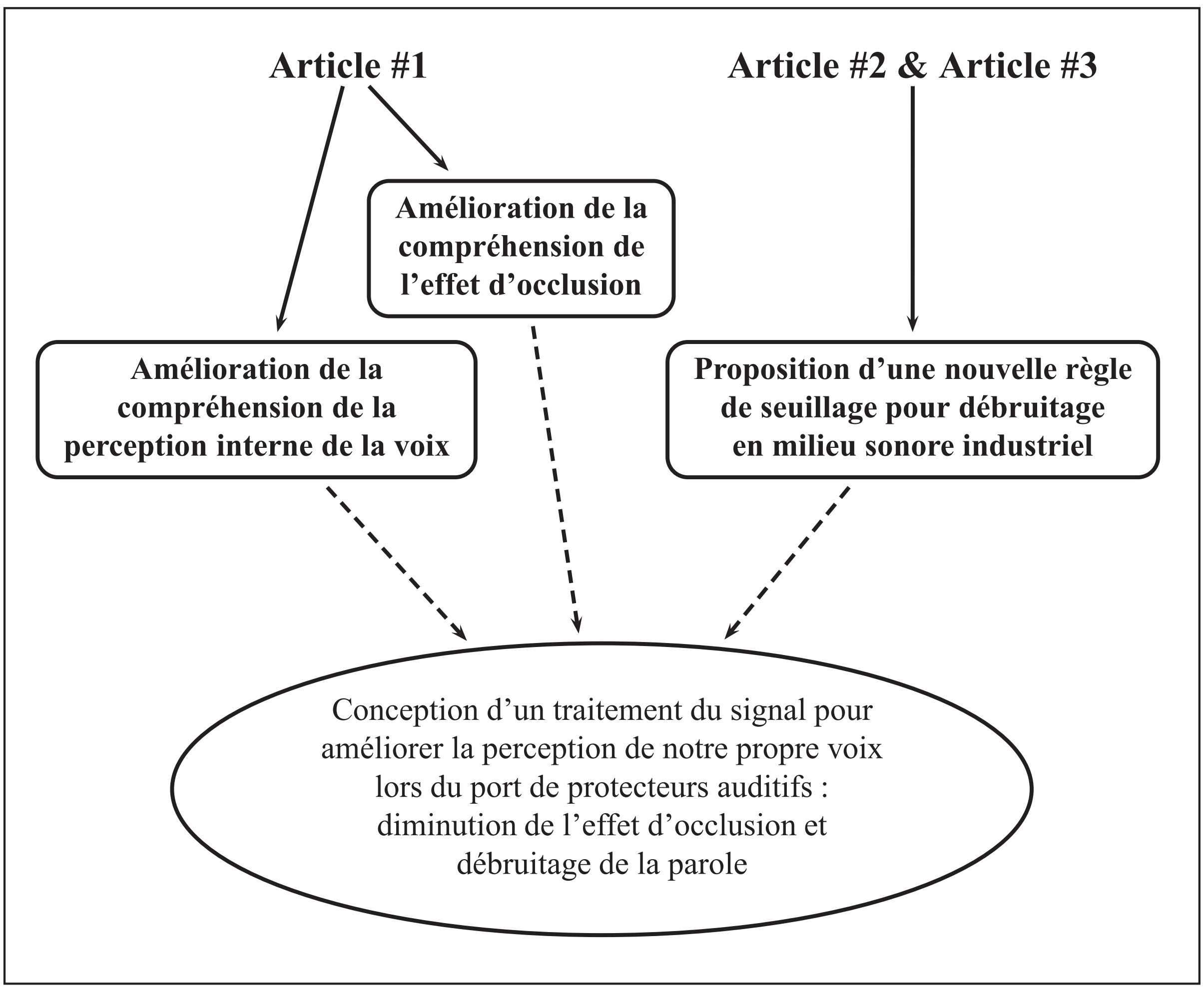

Figure 4.4 Schéma récapitulatif pour la thèse

Dans les trois articles de cette thèse, les avancées majeures obtenues pour la recherche sont :

- Amélioration de la compréhension de l'effet d'occlusion dû à des protecteurs auditifs intra-auriculaires.

- Amélioration de la compréhension de la perception interne de la voix avec et sans protecteurs auditifs intra-auriculaires.
- Proposition d'une nouvelle règle de seuillage pour débruiter la parole par ondelettes en milieu sonore industriel.

Il s'agit à la fois d'avancées dans la recherche fondamentale et dans la recherche appliquée. Les industriels ont ainsi des premières pistes pour améliorer la communication des travailleurs en milieu industriel bruité. Ces avancées sont également les premières pierres qui vont mener à la conception d'un algorithme de traitement du signal qui améliorera la communication en milieu industriel bruité lors du port de protecteurs auditifs, à la fois du point de vue du locuteur et de l'auditeur. Cet algorithme d'une part diminuera l'influence de l'effet d'occlusion pour le locuteur, et d'autre part débruitera la parole pour la réémettre sous le protecteur auditif pour l'auditeur.

BIBLIOGRAPHIE

Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet estimator in nonparametric regression : a comparative simulation study. *Journal of Statistical Software*, 6(6) :1–83.

Ayat, S., Manzuri-Shalmani, M., and Dianat, R. (2006). An improved wavelet-based speech enhancement by using speech signal features. *Computers and Electrical Engineering*, 32(6) :411–425.

Berger, E., Royster, L., Royster, J., Driscoll, D., and Layne, M. (2000). *The Noise Manual*. American Industrial Hygiene Association (AIHA) Press, 5th edition edition.

Berger, E. H. (1986). Hearing protection devices. In Berger, E. H., Ward, W. D., Morrill, J. C., and Royster, L. H., editors, *Noise & Hearing Conservation Manual*, pages 319–382. American Industrial Hygiene Association, Fairfax, Virginia, fourth edition.

Berger, E. H. and Kerivan, J. E. (1983). Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold. *Journal of the Acoustical Society of America*, 74(1) :81–94. 0001-4966 English.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4) :373–384.

Bui, T. and Chen, G. (1998). Translation-invariant denoising using multiwavelets. *Signal Processing, IEEE Transactions on*, 46(12) :3414–3420.

Bárány, E. (1938). A contribution to the physiology of bone conduction. *Acta Oto-Laryngologica*, 26(Suppl.26) :1–223.

Békésy, G. v. (1949). The structure of the middle ear and the hearing of one's own voice by bone conduction. *The Journal of the Acoustical Society of America*, 21(3) :217–232.

Békésy, G. v. (1960a). Bone conduction. In Wever, E. G., editor, *Experiments in hearing*, pages 127–203. McGraw Hill.

Békésy, G. v. (1960b). *Experiments in hearing*. McGraw-Hill, New York.

CSA (2002). Hearing protection devices - performance, selection, care, and use.

Dadson, R. S. and King, J. H. (1952). A determination of the normal threshold of hearing and its relation to the standardization of audiometers. *Journal of laryngology and otology*, 66(8) :366–378.

Dillon, H. (2000). Hearing aid earmolds, earshells and coupling systems. In *Hearing aids*, pages 117–158. Thieme, New York.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432) :1200–1225.

Ehmer, R. H. (1959). Masking patterns of tones. *The Journal of the Acoustical Society of America*, 31(8) :1115–1120.

EN458 (1993). Hearing protectors - recommendations for selection, use, care and maintenance - guidance document. 77.en.

EN458 (1996). Protection. auditive - recommandations pour la sélection, l'utilisation, l'entretien et la maintenance. 81.fr.

Ephraim, Y. (1992). Statistical-model-based speech enhancement systems. *Proceedings of the IEEE*, 80(10) :1526–1555. TY - JOUR.

Fan, G. and Xia, X.-G. (2001). Improved hidden markov models in the wavelet-domain. *Signal Processing, IEEE Transactions on*, 49(1) :115–120.

Flandrin, P. (1998). *Temps-fréquence*. Hermes, 2e edition.

Fodor, I. K. and Chandrika, K. (2003). Denoising through wavelet shrinkage : an empirical study. *Journal of Electronic Imaging*, 12(1) :151–160.

FTQ (1999). De nouveaux outils de formation sur le bruit : Alerte aux décibels ! *Fédération des Travailleurs et Travailleuses du Québec*.

Gao, H.-Y. (1998). Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4) :469–488.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., and Zue, V. (1993). Timit acoustic-phonetic continuous speech corpus.

Gelfand, S. A. (2004). *Hearing : an introduction to psychological and physiological acoustics*. Marcel Dekker, New York, 4th edition. Stanley A. Gelfand. ill. ; 26 cm.

Girard, S. A., Picard, M., Courteau, M., Boisclair, D., Larocque, R., Leroux, T., Turcotte, F., and Simard, M. (2007). Le bruit en milieu de travail : une analyse des coûts poru le régime d'indemnisation. Rapport de recherche, Institut national de santé publique du Québec.

Gustafsson, H., Nordholm, S., and Claesson, I. (2001). Spectral subtraction using reduced delay convolution and adaptive averaging. *Speech and Audio Processing, IEEE Transactions on*, 9(8) :799–807. TY - JOUR.

Hansen, J. H. and Pellom, B. (1998). An effective quality evaluation protocol for speech enhancement algorithms. In *ICSLP-98 : Inter. Conf. on Spoken Language Processing*, volume 7, pages 2819–2822, Sydney, Australia.

Hansen, M. O. (1997). Occlusion effects : Part 1 : Hearing aid users experiences of the occlusion effect compared to the real ear sound level. Technical Report 71, Technical University of Denmark.

Hansen, M. O. (1998). Occlusion effects : Part 2 : A study of the occlusion effect mechanism and the influence of the earmould properties. Technical Report 73, Technical University of Denmark.

Haykin, S. (2002). *Adaptive filter theory*. Prentice-Hall information and system sciences series. Prentice-Hall, Upper Saddle River, N.J., 4th edition. États-Unis.

Howell, P. (1985). Auditory feedback of the voice in singing. In Howell, P., Cross, I., and West, R., editors, *Musical structure and cognition*, pages 259–286. Academic Press, London.

Hétu, R. (1994). Mismatches between audiotry demands and capacities in the industrial work environment. *Audiology*, 33 :1–14.

Industrial Noise Laboratory (2007). Test report hearing protector noise attenuation - ansi s12.6 - 1997 (b) - reat - subject fit. Technical report, Federal University of Santa Catarina.

ISO 389-7, (1996). “Acoustics - reference zero for the calibration of audiometric equipment - part 7 : Reference threshold of hearing under free-fied and diffuse-field listening conditions”.

ISO 4869-1, (1990). “Acoustics - hearing protectors - part 1 : Subjective method for the measurement of sound attenuation”.

ISO 8253-1, (1989). “Acoustics - audiometric test methods - part 1 : Basic pure tone air and bone conduction threshold audiometry”.

ISO 8253-2, (1992). “Acoustics - audiometric test methods - part 2 : Sound field audiometry with pure tone and narrow-band test signals”.

Johnson, D., Papadopoulos, P., Watfa, N., and Takala, J. (2001). Exposure criteria, occupational exposure levels. In *Occupational exposure to noise ; evaluation, prevention and control*. World Health Organization.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society Series B-Methodological*, 59(2) :319–351.

Killion, M. C. (1988). The hollow voice occlusion effect. In *Proceedings of 13th Danavox Symposium*, volume 3, pages 231–241.

Krim, H., Tucker, D., Mallat, S., and Donoho, D. L. (1999). On denoising and best signal representation. *Information Theory, IEEE Transactions on*, 45(7) :2225–2238.

Le Cocq, C., Gargour, C., and Laville, F. (submitted). Wavelet speech enhancement for industrial noise environments. *Speech Communication*.

Lu, C.-T. and Wang, H.-C. (2007). Speech enhancement using hybrid gain factor in critical-band-wavelet-packet transform. *Digital Signal Processing*, 17(1) :172–188.

Lundh, P. (1986). Sound pressure in the ear with vented and unvented earmould. Technical Report 28-8-1, Oticon Electronics A/S.

Neitzel, R. (2009). Occupational noise exposure information for the construction industry. `http://staff.washington.edu/rneitzel/standards.htm`.

NIOSH (1998). Criteria for a recommended standard : occupational noise exposure. Technical Report 98-126, National Institute for Occupational Safety and Health.

NIOSH (2009). Hearing protector device compendium. `http://www2a.cdc.gov/hp-devices/hp_srchpg01.asp`.

NOISEX-92. Institute for Perception-TNO, The Netherlands, Speech Research Unit, RSRE, United Kingdom, 1990.

Nordström, F. (1998). *Time and Frequency Dependent Noise Reduction in Speech Signals.* Master, Lund Institute of Technology.

Nordström, F., Holst, B., and Lindoff, B. (1999). Time and frequency dependent noise reduction in speech signals. In *The International Conference on Signal Processing Applications and Technology*, Orlando, Florida, USA.

Pickett, J. (1956). Effects of vocal force on the intelligibility of speech sounds. *The journal of the acoustical society of america*, 28(5) :902–905.

Reinfeldt, S., Stenfelt, S., Good, T., and Hakansson, B. (2007). Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations. *The Journal of the Acoustical Society of America*, 121(3) :1576–1587.

Robinson, D. W. and Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7 :166–181.

Robinson, G. and Casali, J. G. (2000). Speech communications and signal detection in noise. In Berger, E., Royster, L., Royster, J., Driscoll, D., and Layne, M., editors, *The Noise Manual*. American Industrial Hygiene Association (AIHA) Press, 5th edition edition.

Sheikhzadeh, H. and Abutalebi, H. R. (2001). An improved wavelet-based speech enhancement system. In *Eurospeech 2001*, volume 3, pages 1855–1858, Aalborg, Denmark, Scandinavia.

Small, A. M. J. and Gales, R. S. (1998). Hearing characteristics. In Harris, C. M., editor, *Handbook of acoustical measurements and noise control*, pages 17.1–17.25. Acoustical Society of America, Woodbury, NY, third edition.

Stenfelt, S., Hato, N., and Goode, R. L. (2002). Factors contributing to bone conduction : The middle ear. *The Journal of the Acoustical Society of America*, 111(2) :947–959.

Stenfelt, S. and Reinfeldt, S. (2007). A model of the occlusion effect with bone-conducted stimulation. *International journal of audiology*, 46(10) :595–608.

Stenfelt, S., Wild, T., Hato, N., and Goode, R. L. (2003). Factors contributing to bone conduction : the outer ear. *The Journal of the Acoustical Society of America*, 113(2) :902–913.

Suter, A. H. (1992). The effects of hearing protectors on the perception of speech and warning signals (chapter 3). In *Communication and Job Performance in Noise : A review*, volume 28. ASHA Monograph, Rockville, Maryland, USA, american speech-language-hearing association edition.

Suter, A. H. (2000). Standards and regulations. In Berger, E., Royster, L., Royster, J., Driscoll, D., and Layne, M., editors, *The Noise Manual*. American Industrial Hygiene Association (AIHA) Press, 5th edition edition.

Tantibundhit, C., Boston, J., Li, C., Durrant, J., Shaiman, S., Kovacyk, K., and El-Jaroudi, A. (2007). New signal decomposition method based speech enhancement. *Signal Processing*, 87(11) :2607–2628.

Tonndorf, J. (1972). Bone conduction. In Tobias, J. V., editor, *Foundations of Modern Auditory Theory*, volume 2, pages 195–237. Academic Press, New York.

Tonndorf, J., Greenfield, E. C., and Kaufman, R. S. (1966). The occlusion of the external ear canal : its effect upon bone conduction in cats. *Acta Oto-Laryngologica*, 61(Suppl. 213) :80–104.

Tran Quoc, H. and Hétu, R. (1996). La planification de la signalisation acoustique en milieu industriel : critère de conception des avertisseurs sonores de danger. *Acoustique canadienne*, 24(2) :3–17.

Voix, J. (2006). *Mise au point d'un bouchon d'oreille "intelligent"*. Ph.d. thesis, École de Technologie Supérieure, Montréal, Canada.

Voix, J. and Laville, F. (2009). The objective measurement of individual earplug field performance. *The Journal of the Acoustical Society of America*, in press.

Voix, J., Laville, F., and Zeidan, J. (2002). Filter selection to adapt earplug performances to sound exposure. In *Congrès annuel de l'ACA*, volume 30(3), pages 122–123. CAA-ACA.

Wang, S., Sekey, A., and Gersho, A. (1992). An objective measure for predicting subjective quality of speech coders. *Selected Areas in Communications, IEEE Journal on*, 10(5) :819–829.

Webster, J. C. (1979). Effects of noise on speech. In Harris, M., editor, *Handbook of noise control*, volume chapter 14. McGraw-Hill, New York, 2nd edition edition.

Wegel, R. L. (1932). Physical data and physiology and excitation of the auditoy nerve. *Ann. Otol. Rhinol. Laryngol.*, 41 :740–779.

WHO (1999). Guidelines for community noise. Technical report.

WHO (2001). Occupational exposure to noise : evaluation, prevention and control. Technical report, Federal Institute for Occupational Safety and Health.

Wimmer, V. H. (1986). The occlusion effect from earmoulds. *Hearing Instruments*, 37(12) :19, 57–58.

Yoon, B.-J. and Vaidyanathan, P. (2004). Wavelet-based denoising by customized thresholding. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, volume 2, pages 925–928, Philadelphia, Pennsylvanie.

Yu, S. and Chip-Hong, C. (2007). A generalized time-frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 37(4) :877–889.