Curriculum Vitae

Fathi KALLEL
Nationalité Tunisienne
Marié, 2 enfants
Né le 28.08.1981 à Sfax

Route El ain Km 3 PTT merkez el alia, BP N°43, 3051 Sfax-Tunisie
E-Mail : fathikallel@yahoo.fr
Tél.(mobile) : 21 46 44 75 - Tél.(domicile) : 74 26 87 57

FORMATION

2008 – 2011 – Thèse de doctorat en Génie Electrique et en Electronique, Electrotechnique et Automatique.
- Mention : Très Honorable
- Cotutelle entre l'École Nationale d'Ingénieurs de Sfax, Tunisie et l'Université Claude Bernard de Lyon, France

2005 – 2006 – Master Electronique
- Mention : Bien
- École Nationale d'Ingénieurs de Sfax, Université de Sfax, Tunisie

2002 – 2005 – Cycle de formation d'ingénieurs
- Spécialité : Génie Electrique
- Mention : Bien
- École Nationale d'Ingénieurs de Sfax, Université de Sfax, Tunisie

2001 – 2002 – Classes préparatoires au écoles d'Ingénieurs
- Spécialité : Physique-Chimie
- Institut Préparatoire aux Etudes d'Ingénieurs de Sfax, Université de Sfax, Tunisie

2000 – Baccalauréat
- Spécialité : Sciences Expérimentales
- Mention : Assez bien
- Lycée Technique 9 avril 1938 de Sfax, Tunisie

ACTIVITÉ DE RECHERCHE

Depuis 2012 – Activités post-doctorales
- Développement de nouveaux algorithmes de traitement du signal de parole dans le but d'améliorer l'intelligibilité de la parole chez les sujets porteurs d'une prothèse cocléaire.
- Thème : Traitement du signal de parole, algorithmes de débruitage, prothèse cochléaire.
- Equipes de Recherche : Unité de recherche Advanced Technologies for Medicine and Signals, École Nationale d'Ingénieurs de Sfax, Tunisie & Groupe Parole, Audiologie, Communication & Santé, Centre de Recherche en Neurosciences de Lyon (CNRS UMR5292), Université Claude Bernard de Lyon, France.

2008 – 2011 – Thèse de doctorat en cotutelle
• Titre : Algorithmes de Réduction du Bruit en Vue d'une Amélioration de l'Intelligibilité de la Parole : Cas de la Prothèse Cochléaire
• Thème : Evaluation objective et subjective de différents algorithmes de réduction du bruit (Algorithme de la soustraction spectrale bi-voie, Algorithme de la soustraction interspectrale), Estimation des densités spectrales de puissance des signaux de parole bruités, Estimation des densités spectrales de puissance des bruits, Estimation de la densité interspectrale des bruits, Application : Prothèse cochléaire.
• Equipes de Recherche : Unité de recherche Advanced Technologies for Medicine and Signals, École Nationale dIngénieurs de Sfax, Tunisie & Groupe Parole, Audiologie, Communication & Santé, Centre de Recherche en Neurosciences de Lyon (CNRS UMR5292), Université Claude Bernard de Lyon, France.

2008 – 2011 – Mémoire de Master
• Titre : Implémentation sur DSP TMS320C6416 d'une Stratégie de codage pour Prothèse Cochléaire.
• Thème : Développement et implémentation matérielle de différentes stratégies de codage du signal de parole basées sur les filtres numériques et la transformée de Fourier rapide, Application : Prothèse cochléaire.
• Equipe de recherche : Laboratoire d'Electronique et des Technologies de l'Information, École Nationale dIngénieurs de Sfax, Tunisie.

PUBLICATIONS SCIENTIFIQUES

REVUES INTERNATIONALES

2012

– **F. Kallel**, A. B. Hamida, R. Laboissière et C. Berger-Vachon, 2013. Influence of a Shift in Frequency Distribution and Analysis Rate on Phoneme Intelligibility in Noisy Environment in Simulated Bilateral Cochlear Implant, *Applied Acoustics*, 74 : 10-17.

– **F. Kallel**, M. Ghorbel, M. Frikha, C. Berger-Vachon et A. B. Hamida, 2012. A noise cross psd estimator based on improved minimum statistics method for two-microphone speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :256-264.

– **F. Kallel**, M. Frikha, M. Ghorbel, A. B. Hamida et C. Berger-Vachon, 2012. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :12-20.

– A. Derbel, M. Ghorbel, **F. Kallel**, A. B. Hamida et M. Samet, 2012. Exploring Wavelet Transform Based Methodology for Cochlear Prosthesis Advanced Speech Processing Strategy, soumis au journal Acta Acoustica.

2008

– A. Derbel, **F. Kallel**, M. Samet et A. B. Hamida, 2008. Bionic wavelet transform based on speech processing dedicated to a fully programmable stimulation strategy for cochlear prostheses. *Asian Journal of Scientific Research*, 1 :293-309.

Conférences Internationales

2010

– **F. Kallel**, A. Jeanvoine, A. B. Hamida et C. Berger-Vachon, 2010. Etude de l'effet du mode de stimulation sur l'intelligibilité de la parole en milieu silencieux et en milieu bruité. *Handicap 2010*, Porte des Versailles, Paris-France, Juin 2010.

2009 – **F. Kallel**, D. Daoud, M. Ghorbel, et A. B. Hamida, 2009. Comparaison des différents algorithmes de débruitage du signal de parole pour les aides auditives binaurales. 5^{th} *International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, Hammamet-Tunisie, March 2009.

– D. Daoud, **F. Kallel**, M. Ghorbel, et A. B. Hamida, 2009. Spatial filtering based speech enhancement for binaural hearing aid. 6^{th} *International Multi-Conference on Systems, Signals and Devices*, Djerba-Tunisie, March 2009.

2007 – A. Derbel, **F. Kallel**, A. B. Hamida et N. Ellouze, 2007. Wavelet-Based Parameterisation of Speech Signal Dedicated to Cochlear Prosthesis", 5^{th} *International Multi-Conference on Systems, Signals and Devices*, Hmmamet-Tunisie, March 2007.

– A. Derbel, **F. Kallel**, A.B. Hamida, 2007. Wavelet Filtering Based on Mellin Transform Dedicated to Cochlear Prostheses. 29^{th} *IEEE International Conference on Engineering in Medicine and Biology Society*, Lyon-France, Octobre 2007.

Conférences Nationales

2007 – **F. Kallel**, A. B. Hamida, 2007. Implémentation sur DSP d'une Stratégie de Stimulation Flexible pour Prothèse Cochléaire Basée sur un Banc de Filtres. *Septièmes Journées Scientifique en Génie Electrique et Informatique (GEI'07)*, Monastir-Tunisie.

– A. Derbel, **F. Kallel**, A. B. Hamida, N. Ellouze, 2007. Conception et Implémentation d'une Stratégie de Stimulation Basée sur la Transformée en Ondelette pour les Prothèses Cochléaires. *Septièmes Journées Scientifique en Génie Electrique et Informatique (GEI'07)*, Monastir-Tunisie.

COMPÉTENCE

Scientifiques – Electronique analogique, Electronique numérique, Traitement analogique et numérique du signal, Traitement d'images, Techniques multimédia, Appareillage Biomédical (Prothèse auditive, Prothèses Cochléaire, Audiométrie, Antidouleur, Dialyse...)

Langages de programmation – C/C++, Pascal, SPICE, ModelSim

Logiciels – Matlab, Visual Basic, Visual C++, R, Code Composer Studio

Cibles – PIC, Processeur de traitement du signal "DSP"(TMS320C6416)

Bureautique – Word, Excel, Powerpoint, LaTeX

ACTIVITÉS PÉDAGOGIQUES

2010 – Participation au journées de formation sur les Processeurs de Traitement de Signal DSP "TMS320C64x", Sousse-Tunisie.

2008 – Participation au premières journées de l'école doctorale Sciences et Technologie, Monastir-Tunisie.

2006 – Accomplir la formation English for Research Purposes organisée dans le cadre de la formation continue en anglais à l'ENIS.

– Accomplir le cycle de formation pédagogique organisé dans le cadre des formations de Mastère à l'ENIS.

Avril 2005 – Participation à l'université de printemps de pédagogie, ENIS.

ACTIVITÉS D'ENSEIGNEMENT

2012 – 2013 – Maître Assistant en Traitement du Signal et de l'image à l'Institut Supérieur d'Informatique et des Mathématiques de Monastir.
- Cours en architecture des processeurs de traitement du signal ; LA STIC 3.
- Cours et TD en programmation des processeurs de traitement du signal ; LA STIC 3 et LF3 STIC .
- Cours en processeurs de traitement du signal ; 3ème année Cycle ingénieur.
- Cours en traitement numérique du signal ; 1ère année Master recherche en Electronique.

2011 – 2012 – Assistant à l'Institut Supérieur d'Informatique et du Multimédia de Gabès.
- 21 heures de cours en codes correcteurs et traitement du signal ; Mastère Pro.SSI.
- 42 heures TP en codes correcteurs et traitement du signal ; Mastère Pro.SSI.
- 42 heures de cours en architecture des ordinateurs ; LATMW1.
- 42 heures de TD en architecture des ordinateurs ; LATMW1.
- 21 heures de cours en processeurs dédiés ; Mastère Pro. en Systèmes Embarqués.
- 42 heures de TP en processeurs dédiés ; Mastère Pro. en Systèmes Embarqués.

2010 – 2011 – Assistant contractuel à la faculté des sciences de Gabès.
- 22.5 heures de cours en fondement du multimédia ; LARI2.
- 22.5 heures de TD en fondement du multimédia ; LARI2.
- 11 heures de cours en fondement du multimédia ; LFSI3.
- 22.5 heures de cours en vision industrielle ; LATIM1.
- 22.5 heures de cours en analyse des images numériques ; LATIM3.
- 22.5 heures de cours en logiciels mathématiques ; LARI1.
- 15 heures de TP en traitement du signal ; LARI1.

2009 – 2010 – Assistant contractuel à la faculté des sciences de Gabès.
- 22.5 heures de cours en transmission du signal ; LATIM2.
- 11 heures de cours en fondement du multimédia ; LATIM1.
- 60 heures de TP en fondement multimédia ; LATIM1.
- 30 heures de TP en traitement d'images ; LATIM3.
- 22.5 heures de cours en traitement du signal ; LATIM 1.
- 60 heures de TD en traitement du signal ; LATIM1.
- 11 heures de TD réseaux sans fils ; LATIM1.

2008 – 2009 – Assistant contractuel à la faculté des sciences de Gabès.
- 11 heures de cours en fondement du multimédia ; LATIM1.
- 60 heures de TD en traitement du signal ; I3.
- 120 heures de TD en architecture des ordinateurs ; LATIM3.
- 22.5 heures de cours en codage des objets multimédias ; LATIM1.
- 15 heures de TD en codage des objets multimédias ; LATIM1.

2007 – 2008 – Assistant contractuel à la faculté des sciences de Gabès.
- 11 heures de cours en fondement du multimédia ; LATIM1.
- 45 heures de TD en algorithmique et structure de données I ; LFSI1.
- 45 heures de TD en algorithmique et structure de données II ; LFSI1.
- 90 heures de TP en traitement d'images ; I4.
- 90 heures de TP en fondement du multimédia ; LATIM1.
- 90 heures de TD en traitement du signal ; LATIM1.

2006 – 2007 – Assistant contractuel à l'Institut Supérieur d'Informatique et de Multimédia de Sfax.
- 45 heures de TD en système informatique ; TIM1.
- 60 heures de TP en système informatique ; TIM1.
- 120 heures de TP en programmation structuré en langage assembleur ; TMSI2.

2005 – 2006 – Contrat Etudiant Chercheur à l'Institut Supérieur de Biotechnologie de Sfax.
- 60 heures de TP en instrumentation biomédicale ; IBM3.
- 100 heures de TP en bureautique ; IBM1.
- 40 heures de TP en langage C++ ; IBM2.
- 60 heures de TP en programmation en langage Matlab ; IBM2.
- Participation à des activités de travaux pratique, de mini projets et d'encadrement en traitement de signal, GE2, option ENT,ENI.

TRAVAUX D'ENCADREMENT

Participation à l'encadrement de projets de fin d'études

- 6 projets de fin d'études, Génie Electrique, Ecole Nationale d'Ingénieurs de Sfax.
- 3 projets de fin d'études, Génie Electrique, Ecole Nationale d'Ingénieurs de Sfax. et Telnet-Sfax.
- 2 projets de fin d'études, Licence en Informatique, Faculté des Sciences de Sfax.

Encadrement de projets de fin d'études

- 2 projets de fin d'études, LA3 STIC, ISIM-Monastir.
- 2 projets de fin d'études, ING 3 Electronique, ISIM-Monastir.
- 3 projets de fin d'études, Licence Fondamentale en Sciences Informatique, Faculté des Sciences de Gabès
- 2 projets de fin d'études, Instrumentation Biomédicale, Institut Supérieur de Biotechnologie de Sfax et CHU de Sfax

ACTIVITÉS DIVERSES

Contrats de recherche

2008 – 2011 – Membre du projet scientifique de recherche tunisio-français CMCU (code : 09G/1421), intitulé : Étude et implémentation d'algorithmes de débruitage du signal de la parole pour la réhabilitation des surdités.

– Membre du projet scientifique de recherche tunisio-français DGRS/CNRS (code : 09/R11-18), intitulé : Étude et implémentation d'algorithmes de débruitage du signal de parole pour implants binauraux.

Stages & *Séjours scientifique*

Août 2003 – Stage technicien à la Société Tunisienne d'Electricité et du Gaz "STEG" de Sfax .

Août 2004 – Stage ouvrier à Tunisie Télécom-Centre de Transmission de Données de Sfax.

2008 – 2011 – Stages dans le cadre de la cotutelle de Thèse, Université Claude Bernard-Lyon1, sous la direction du Pr Christian BERGER VACHON.

Decembre 2012 – Séjours scientifique de haut niveau SSHN, Université Claude Bernard-Lyon1.

L A N G U E S

Arabe – Langue Maternelle, lu parlé et écrit courant

Français – Lu, parlé et écrit

Anglais – Lu, parlé et écrit

Résumé

La surdité ou bien le dysfonctionnement du système auditif est un handicap qui peut être parfois grave pour l'être humain. Cette surdité conduit le malentendant à vivre dans un monde de silence, isolé de toute vie sociale. Les avancées technologiques en ingénierie biomédicale associées aux développements de la médecine ont permis une amélioration globale des conditions de vie et une augmentation sensible de la longévité de l'homme en donnant naissance à de nouveaux appareillages biomédicaux tel que la prothèse cochléaire. En effet, la prothèse cochléaire est un appareillage biomédical, implantable au niveau de l'oreille humaine, qui permet de faire bénéficier certaines personnes atteintes d'une surdité profonde ou totale bilatérale d'un niveau d'audition inaccessible avec les prothèses auditives traditionnelles.

La prothèse cochléaire assure la stimulation directe des neurones cochléaires. La pose d'électrodes dans des zones bien définies dans la cochlée permet de stimuler sélectivement les cellules sensorielles suivant différentes fréquences et différentes intensités électriques qui sont générées suite à un traitement spécifique du signal de parole. La stimulation électrique apportée par les électrodes permet un niveau de compréhension intéressant tout en notant une adaptation progressive du patient à son appareillage suite à une phase de rééducation.

Différents travaux de recherche ont été établis afin d'évaluer l'intelligibilité de la parole chez les sujets implantés en environnements silencieux et bruité. Les résultats ont montré une bonne intelligibilité de la parole variant entre 80% et 90% en milieu silencieux. Toutefois, les capacités de perception de la parole par les patients implantés se dégradent en environnement bruité. Afin d'améliorer l'intelligibilité de la parole en milieu bruité, différents algorithmes de réduction de bruit, appelés aussi algorithmes de débruitage, ont été développés dans le cas de l'implant cochléaire. Ces algorithmes peuvent être classés principalement en deux catégories. Pour la première catégorie, les algorithmes de débruitage sont intégrés en totalité au niveau de la stratégie de codage adopté pour l'analyse du signal de parole au niveau de l'implant en modifiant certaine fonction tel que la fonction de compression. Pour la deuxième catégorie, une étape de prétraitement basés sur un algorithme de débruitage spécifique est tout d'abord adoptée. Le signal de parole rehaussé ainsi obtenu à la suite de cette étape de prétraitement est ensuite traité par l'algorithme de codage utilisé par l'implant. Dans ce cas, différents algorithmes de débruitage initialement développés pour des sujets normoentendants ont été adoptés dans le cas de l'implant cochléaire.

Le développement progressif de la technologie a rendu possible l'implantation des patients d'une manière bilatérale. L'implantation cochléaire bilatérale consiste à poser un implant cochléaire au niveau de chaque oreille. Elle permet de plus en plus l'accès à l'audition humaine naturelle qui est une audition binaurale. Ceci est assuré grâce à la restauration des capacités de localisation spatiale, l'amélioration de la sélectivité fréquentielle et de la reconnaissance de la parole dans le bruit. De nombreux travaux de recherche ont mis en évidence les performances d'une stimulation bilatérale par rapport à une stimulation unilatérale en milieux silencieux et bruité. Les résultats ont montré qu'une stimulation bilatérale permet d'améliorer l'intelligibilité de la parole surtout en milieu bruité par rapport à une stimulation unilatérale. D'autre part, le patient implanté se trouve très souvent dans des environnements où le niveau de bruit est assez élevé. Dans ce cas, même avec une stimulation bilatérale, l'intelligibilité de la parole reste insuffisante et la perception sonore est difficile. Afin d'améliorer davantage l'intelligibilité en milieu bruité, différents algorithmes de débruitage ont été développés dans le cas de l'implant cochléaire bilatéral. Ces algorithmes, basés sur un traitement multi-microphones, présentent l'inconvénient du coût assez élevé et de la complexité. Un traitement à base de deux microphones (chaque implant est équipé d'un seul microphone) pourrait être considéré comme étant le meilleur compromis entre un traitement mono-microphone et un traitement multi-microphones.

Dans ce travail de thèse, nous avons traité le problème de l'intelligibilité de la parole dans le cas de l'implant cochléaire, particulièrement l'implant cochléaire bilatéral, en proposant de nouvelles approches de traitement du signal pour le débruitage du signal de parole. Ce travail est structuré principalement en trois parties :

Dans la première partie, nous avons présenté une première approche de traitement du signal de parole visant l'amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire. Un protocole expérimental a été adopté pour la comparaison des performances de trois modes de stimulation cochléaire. Les performances de ces trois modes de stimulation ont été comparées dans le cas de deux vitesses d'analyse, en milieux silencieux et bruité à différents niveaux du RSB. Le premier mode est la stimulation unilatérale. Le deuxième mode est la stimulation bilatérale symétrique où les bancs de filtres utilisés pour l'analyse des signaux de parole au niveau des voies droite et gauche sont identiques. Le troisième mode est la stimulation bilatérale décalée où les deux oreilles sont stimulées par des signaux différents (le banc de filtres utilisé pour l'analyse du signal de parole au niveau de la voie droite est fréquentiellement décalé par rapport à celui de la voie droite). Les performances de ces trois modes de stimulation ont été comparées d'une manière subjective avec une population de cinquante sujets normoentendants.

Dans la deuxième partie, les performances de deux algorithmes bi-voie pour la réduction du bruit dédiés pour implant cochléaire bilatéral ont été comparées. Ces algorithmes sont basés sur deux étapes de traitement : Une étape d'estimation de la densité spectrale de puissance (dsp) du bruit au niveau de chaque voie suivi d'une étape d'estimation des signaux rehaussés. Le principe d'un algorithme bi-voie pour l'estimation de la dsp du bruit est d'abord présenté. Cet algorithme a été basé sur le calcul des dsp et de la densité interspectrale (dip) des signaux bruités et il a été développé sous l'hypothèse des bruits parfaitement décorrélés. L'estimation des signaux rehaussés a été déterminée en se basant sur la technique de la soustraction spectrale. Le principe des algorithmes de la soustraction spectrale non linéaire bi-voie et la soustraction spectrale multi-bande bi-voie sont alors détaillés. Les performances de ces deux algorithmes ont été évaluées et comparées d'une manière subjective dans le cas de l'implant cochléaire bilatéral, en simulation en premier lieu avec des sujets normoentendants puis avec des sujets implantés portant l'implant cochléaire Digisonic SP binaural de Neurelec. Cette étude comparative est établie à différents niveaux du RSB et dans le cas de deux configurations spatiales des sources de bruit (d'abord en présence d'une seule source de bruit, puis en présence de trois sources de bruit).

Enfin, dans la troisième partie, un deuxième algorithme de réduction de bruit dédié pour l'implant cochléaire bilatéral a été proposé. Cet algorithme est basé sur la méthode de la soustraction interspectrale qui est développé sous l'hypothèse des bruits peu corrélés ou diffus. L'exploitation de l'algorithme de la soustraction interspectrale repose principalement sur l'estimation de la dip des bruits. Dans ce travail, une nouvelle approche basée sur l'algorithme des statistiques minimales améliorées a été proposée pour l'estimation de la dip des bruits. Deux autres approches présentées dans la littérature basées sur la technique de détection d'activité vocale et la technique des statistiques minimales ont été aussi considérées. Les performances de l'algorithme de la soustraction interspectrale sont évaluées dans le cadre de l'implant cochléaire bilatéral en simulation dans le cas de chacun des estimateurs de la dip des bruits considérés avec des sujets normo-entendants. Cette étude comparative est établie dans les mêmes conditions expérimentales que l'expérience précédente (différents niveaux du RSB et deux configurations des sources de bruit).

Attestation

Le travail de M.Fathi Kallel a été effectué dans le cadre d'une co-tutellle entre l'Ecole Nationale d'Ingénieurs de Sfax et l'université Claude-Bernard de Lyon dans le cadre d'un projet Franco-Tunisien Egide.

L'étude a porté sur le codage du signal acoustique en vue de l'application à l'implant cochléaire et elle était intitulée « Algorithmes de réduction du Bruit en vue de l'amélioration de l'intelligibilité de la Parole ; cas de la prothèse cochléaire ». Elle a donné lieu à trois articles qui ont été publiés par la revue « Applied Acoustics ».

La soutenance a eu lieu en Tunisie, le 13 décembre 2011, devant un jury comportant deux professeurs et un Industriel français et quatre professeurs tunisiens. Elee a permis de mettre en évidence la qualité du travail, ce qui a conduit le jury à féliciter le candidat et à lui attribuer la mention maximale autorisée. Compte tenu de ses travaux, Fathi Kallel a été recruté dans l'enseignement supérieur en Tunisie (Université de Gabès).

Dans ces conditions, je soutiens sans réserve la candidature du travail de M.Fathi Kallel pour le prix de thèse de l'IFRATH.

Christian Berger-Vachon

Professeur

Université Lyon 1

PARIS 8 UNIVERSITE
Vincennes - Saint Denis

UFR TES, Département Santé, Ingénierie, Handicap

Paris Novembre 2011

HANDI MASTER
Pr. Jaime López Krahe

Titre : *«Algorithmes de Réduction du Bruit en vue d'une Amélioration de l'Intelligibilité de la Parole : Cas de la Prothèse Cochléaire»*
Auteur Fati Kallel
Rapport de thèse en co-tutelle présenté pour l'obtention du grade de Docteur de l'université Lyon 1 (France) et de l'Université de Sfax (République Tunisienne)
Date de soutenance : 13 décembre 2011
Codirecteurs : Ahmed Ben Hamida (Sfax), Christian Berger Vachon(Lyon I)

Rapport de thèse

Ce travail de thèse fait suite à une collaboration entre le LETI (ENIS, Université de Sfax Tunisie) et le Laboratoire Neuroscience de l'Université Claude Bernard, Lyon (France), dans le cadre de travaux de recherches scientifiques. Cette thèse est donc le résultat d'une cotutelle international suite à cette collaboration.

Le rapport est organisé en

- une introduction générale de présentation du sujet,
- cinq chapitres :
  - Implant cochléaire et effet de bruit
  - Techniques de base pour la réduction de bruit
  - Effet de la distribution fréquentielle sur l'intelligibilité de la parole en milieu silencieux et bruité : Cas de l'implant cochléaire bilatéral
  - Exploitation de la technique de soustraction spectrale bi-voie pour la réduction du bruit ; Cas de l'implant cochléaire bilatéral
  - Exploitation de la technique de soustraction inter spectrale pour la réduction de bruit. Cas de l'implant cochléaire bilatéral
- une conclusion, une bibliographie assez conséquente et une annexe.

Le document complet comporte 127 pages.

Le travail est orienté à la réduction du bruit de parole dans un contexte de prothèses et en particulier des implants cochléaires au moyen de divers algorithmes spécifiques développés dans le cadre de ce travail. Il a étudié expérimentalement et comparé l'effet de ces algorithmes de débruitage sur la reconnaissance de la parole chez des implantés cochléaires principalement bilatéraux.

Nous sommes devant un rapport pour l'obtention du grade de Docteur, qui décrit un travail originale et qui s'attaque à un problème intéressante dans un contexte de recherche en traitement de signal appliqué au handicap auditif. Le travail présenté est important et est appuyé par de publications internationales de très bon niveau.

En conclusion, je considère que le travail de M. Fathi Kallel, présenté dans le document intitulé *«Algorithmes de Réduction du Bruit en vue d'une Amélioration de l'Intelligibilité de la Parole : Cas de la Prothèse Cochléaire»,* ainsi que la sélection des travaux personnels et sa production scientifique, est concorde avec un travail correspondant à celui demandé pour l'obtention du grade de docteur dans une université française. Par conséquent je donne un avis très favorable pour que cette thèse soit soutenue publiquement.

Fait à Paris le 10 novembre 2011

Jaime López Krahe, professeur

THIM, D-303, 2, rue de la Liberté, 93 526 Saint Denis Cedex - : Tél. : 01 49 40 73 47 - Fax : 01 49 40 73 48
jlk@univ-paris8.fr ; www.master-handi.fr APE 8541Z SIRET 199 318 270 00014

المدرسة الوطنية للمهندسين بتونس

école nationale d'ingénieurs de Tunis

Raport sur la Thèse de Doctorat de monsieur Fathi Kallel

Le travail de thèse de monsieur Fathi Kallel est intitulée Algorithmes de Réduction du Bruit en vue d'une Amelioration de l'Intelligibilite de la Parole : Cas de la Prothèse Cochleaire. Ce travail a été réalisé dans le cadre d'une thèse en cotutelle entre l'Ecole Nationale d'Ingénieurs de Sfax et l'Université Claude Bernard Lyon 1.

Ce travail porte sur l'étude de l'effet des algorithmes de réduction de bruit en reconnaissance vocale pour des appareillages de prothèses cochléaires et plus spécifiquement des implants bilatéraux. Il concerne dans un premier temps l'effet du mode de stimulation et de la vitesse de stimulation en environnements silencieux et à différents niveaux de bruit, sur les taux de reconnaissance des phonèmes. Un simulateur d'implant basé sur un vocodeur à canaux permet de modéliser le fonctionnement de l'implant cochléaire. Les signaux traités sont présentés à des sujets sains pour trois modes de stimulation: une stimulation bilatérale symétrique, une stimulation bilatérale décalée et une stimulation unilatérale et pour deux vitesses de stimulation à 250 et à 500 par seconde. Les tests sont opérés en environnements silencieux et bruité à 6dB, 0dB et -6dB. Les résultats statistiques montrent en milieu silencieux de meilleures performances dans le cas d'une stimulation bilatérale, et pas de différence significative entre les modes symétrique et décalé. La vitesse de stimulation n'a pas d'effet dans ce cas. En milieu bruité, le mode de stimulation bilatéral décalé présente des performances supérieures aux autres modes. Dans un deuxième temps, l'étude porte sur les performances des algorithmes de réduction de bruit basés sur l'algorithme classique de la soustraction spectrale, et l'algorithme de soustraction spectrale multi-bande. Les performances des algorithmes sont testées d'une manière subjective sur des sujets sains dans le cas d'un bruit blanc et dans le cas d'un bruit cocktail Party de deux locuteurs. Les résultats montrent que l'estimation bi-voies du bruit présente des performances nettement supérieures à l'estimation mono-voie. Les meilleurs résultats sont obtenus par la méthode de soustraction spectrale multi-bande. Dans un troisième temps l'étude a porté sur les performances de la soustraction interspectrale comparativement aux techniques de soustraction spectrale. Une estimation bi-voies du bruit est considérée en présence des deux types de bruit bruit blanc et bruit de type cocktail Party. Les résultats n'ont pas montré d'effet significatif de la technique de réduction de bruit.

Le rapport est structuré comme suit :

Le premier chapitre introduit la notion de bruit et son influence sur l'intelligibilité de la parole.
Le second chapitre est consacré à l'étude des principaux algorithmes de réduction du bruit, et l'estimation de la densité spectrale de puissance du bruit. En particulier la soustraction spectrale, la soustraction spectrale multi-bande, et le filtrage de Wiener.
Le troisième chapitre, porte sur une étude comparative des performances de trois modes de stimulation cochléaire dans les cas de deux vitesses d'analyse en milieux silencieux et bruité à différents niveaux du RSB.
Le quatrième chapitre traite des performances des algorithmes de réduction de bruit bi-voie dédiés pour implant cochléaire bilatéral. Les algorithmes basés sur les techniques de la soustraction spectrale non linéaire bi-voie et la soustraction spectrale multi-bande bi-voie sont utilisés pour l'estimation des signaux rehaussés. Les performances des deux algorithmes sont évaluées et comparés dans le cas de l'implant cochléaire bilatéral, en simulation en premier lieu avec des sujets sains et sur des sujets implantés portant l'implant cochléaire DIGISONIC SP binaural de Neurelec. L'étude est opérée à différents niveaux de RSB et dans le cas de deux configurations spatiales des sources de bruit.

Dans le cinquième et dernier chapitre, porte sur un algorithme de réduction de bruit, dédié pour l'implant cochléaire bilatéral, et basé sur la méthode de la soustraction inter-spectrale. Cet algorithme est développé sous l'hypothèse des bruits peu corrélés ou diffus. Les performances de l'algorithme de soustraction inter-spectrale sont comparées dans le cadre de l'implant cochléaire bilatéral en simulation. Les résultats ont montré que l'algorithme de la soustraction inter-spectrale permet d'améliorer significativement l'intelligibilité de la parole chez les sujets bilatéralement implantés avec une supériorité pour l'algorithme basé sur la technique des statistiques minimales améliorées.
Ce travail a fait l'objet de deux publications acceptées dans Applied Acoustics, et un troisième article soumis au même journal.

A ce titre je donne un avis favorable à la soutenance de la thèse en génie électrique de l'Ecole Nationale d'Ingénieurs de Sfax, de monsieur Fathi Kallel.

Professeur Noureddine Ellouze

Université de Tunis Elmanar

Tunis le 31 Septembre 2011

Université Claude Bernard Lyon 1

Division des Etudes et de la Vie Universitaire

Service des Etudes Doctorales

Bâtiment l'Atrium
43, Bd du 11 Novembre 1918
69622 VILLEURBANNE cedex

UNIVERSITÉ DE LYON

N° ordre : 293 – 2011

RAPPORT D'APRES SOUTENANCE [(1)]

Relatif à la thèse soutenue
en cotutelle avec l'Ecole Nationale d'Ingénieurs de Sfax – Université de Sfax - Tunisie
par Monsieur Fathi KALLEL
le 13 décembre 2011

Le candidat a présenté un exposé clair et concis qui montre ses qualités pédagogiques et la maîtrise de son domaine de recherche portant sur la réduction du bruit pour l'amélioration de l'intelligibilité de la parole dans le domaine de l'implant cochléaire.

Le jury a apprécié les réponses du candidat aux multiples questions, pointues, qui lui ont été posées.

Le jury unanime reconnait le candidat digne du titre de docteur en Génie Electrique de l'Ecole Nationale Supérieure de Sfax (Tunisie) et en EEA de l'Université Claude-Bernard de Lyon (France).

L'université Lyon 1 ne délivre pas de mention en application de l'article 20 de l'arrêté du 7 août 2006 relatif à la formation doctorale et de la décision du CA du 21 juin 2011

Ahmed BEN HAMIDA	Christian BERGER-VACHON	Noureddine ELLOUZE	Lotfi KAMMOUN	Jaime LOPEZ KRAHE

(1) Attention : Ce document doit être signé en séance par tous les membres du jury.

Applied Acoustics 73 (2012) 12–20

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant

Fathi Kallel [a,b,*], Mondher Frikha [a], Mohamed Ghorbel [a], Ahmed Ben Hamida [a], Christian Berger-Vachon [b]

[a] Research Unit in Advances Technologies for Medical and Signals (ATMS), Laboratory of Electronics and Information Technologies (LETI), National Engineering School of Sfax, University of Sfax, Route Soukra km 3, Sfax, B.P.W, 3038, Tunisia

[b] PACS Team, INSERM Unit 1028: "Cognition and Brain Dynamics", Lyon Neurosciences Centre, EPU-ISTIL, Claude Bernard University, Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France

ARTICLE INFO

Article history:
Received 11 October 2010
Received in revised form 19 May 2011
Accepted 23 June 2011
Available online 22 July 2011

Keywords:
Dual-channel noise PSD estimation
DC-NLSS
DC-MBSS
BCI

ABSTRACT

In this paper, two speech enhancement algorithms (SEAs) based on spectral subtraction (SS) principle have been evaluated for bilateral cochlear implant (BCI) users. Specifically, dual-channel noise power spectral estimation algorithm using power spectral densities (PSD) and cross power spectral density (CPSD) of the observed signals was studied. The enhanced speech signals were obtained using either Dual Channel Non Linear Spectral Subtraction 'DC-NLSS' or Dual-Channel Multi-Band Spectral Subtraction 'DC-MBSS' algorithms. For performance evaluation, some objective speech assessment tests relying on Perceptual Evaluation of Speech Quality (PESQ) score and speech Itakura-Saito (IS) distortion measurement were performed to fix the optimal number of frequency band needed in DC-MBSS algorithm. In order to evaluate the speech intelligibility, subjective listening tests were assessed with 50 normal hearing listeners using a specific BCI simulator and with three deafened BCI patients. Experimental results, obtained using French Lafon database corrupted by an additive babble noise at different Signal-to-Noise Ratios (SNR), showed that DC-MBSS algorithm improves speech understanding better than DC-NLSS algorithm for single and multiple interfering noise sources.

1. Introduction

Most cochlear implant (CI) users perform well in quiet listening conditions and many users can now achieve even more than 80% word recognition scores regardless the used device [28]. However, speech recognition scores are enormously degraded in noisy environments [25]. Furthermore, as mentioned by CI users, better and comfortable speech recognition in noisy environments would be considered as one of their most significant challenges [23]. As a first trial to improve speech intelligibility in noisy environments, individuals with severe to profound hearing loss can now be implanted with two cochlear implants, one in each ear. In fact, a bilateral cochlear implantation provides patients the advantages of bilateral information. Bilateral hearing permits optimal performance of the auditory system, with a better understanding of speech in quiet and even better understanding in noisy environments [22]. Recent works compared also speech performance in noisy environment with matched bilateral CI with respect to unilateral CI users. BCI group showed significantly better performance on speech perception in noisy environments compared to the unilateral CI subjects [26,7,8]. Different other clinical studies have demonstrated a substantial increase in speech intelligibility with bilateral cochlear implants compared to monaural listening configurations in noise [30,33,18].

To reduce background effects of noise, some speech enhancement algorithms originally developed for normal hearing listeners have been applied to CI speech processing [17,36,14,11]. These algorithms were able to somewhat improve CI users' performance in noisy listening conditions. Considerably, larger benefits in speech intelligibility could be obtained when resorting to multi-microphone adaptive signal processing strategies, instead. Such strategies make use of spatial information due to the relative position of the emanating sounds, and could therefore better exploit situations in which the target and masker are spatially separated [13,34,5]. Several noise-reduction algorithms using two or more microphones were also available, and most of these proposed algorithms were based on beamformer techniques, especially, the adaptive beamformer algorithms. The performance of adaptive beamforming with two microphones with bilateral cochlear implant was assessed by different studies [29]. In the study of Chung et al. [5], authors conducted experiments to investigate whether directional microphones and adaptive multi-channel noise reduction algorithms could enhance overall CI performance. Results indicated that directional microphones could provide an average improvement of around 3.5 dB. Spriet et al. [29] investigated the

* Corresponding author.
E-mail address: fathikallel@yahoo.fr (F. Kallel).

0003-682X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2011.06.010

performance of the beam pre-processing strategy in the Nucleus Freedom speech processor with five CI users. The performance with the beam strategy was evaluated at two noise levels and with two types of noise, speech-weighted noise and multi-talker babble. The tested algorithm improved the speech reception threshold by approximately 5 to 8 dB. Kokkinakis and Loizou [16] proposed a multi-microphone based adaptive noise reduction strategy exploiting information simultaneously collected by two behind-the-ear processors (BTE) microphones. Four microphones were employed (two omni-directional and two directional) in each of the two BTE processors (one per ear). Results indicated that the proposed multi-microphone strategies improved speech understanding in single and multi-noise source scenarios.

We note that mush of the focus of the previously published studies has been to investigate, in the mono-channel case, the pre-processing noisy speech signal by noise reduction algorithms in order to feed enhanced signals to CI listeners. Only a small number of SEAs have been evaluated to the bilateral case. As one contribution of this work is the application of SEAs for the BCI users in an uncorrelated additive babble noise environments. We propose then new SEAs built upon series of previously published works on spectral subtraction algorithms. The first is based on a non linear spectral subtraction approach according to the work of Berouti et al. [1]. Whereas the second is built on a multi-band spectral subtraction algorithm proposed by Kamath and Loizou [15] and Udrea et al. [32].

Our proposed SEAs namely DC-NLSS and DC-MBSS are an extension of the previously mentioned mono-channel spectral subtraction algorithms to the dual-channel conception case. However, spectral subtraction principles are combined together with a noise PSD estimation technique based on the use of two omni-directional microphones. This noise PSD estimator takes into account the coherence between both received noisy speech signals.

The paper is outlined as follows. Section 2 provides theoretical overview of SEAs. Section 3 derives bilateral cochlear implant simulator principle. Section 4 evaluates the experimental results. Section 5 gives an overall discussion of all obtained results. Finally, Section 6 devotes to the conclusion.

2. Speech enhancement algorithms

The bilateral cochlear implant configuration is as illustrated in Fig. 1. Both left and right cochlear implants are fitted with one microphone. The two received noisy speech signals ($y_1(n)$ and $y_2(n)$) are processed together with the proposed SEAs to generate enhanced speech signals ($\hat{s}_1(n)$ and $\hat{s}_2(n)$) which are then used for stimulation.

We suppose that the noise received by the microphones can be represented by two additive uncorrelated babble noise signals so that the picked up noisy speech signals can be expressed in temporal domain as follows:

$$y_i(k) = s_i(k) + d_i(k) \quad i = \{1, 2\} \tag{1}$$

where 'i' is the microphone index, $y_i(k)$, $s_i(k)$ and $d_i(k)$ represent respectively noisy speech, clean speech and noise signals.

Note that $i = 1$ corresponds to the signal picked up by the microphone placed in the right ear and $i = 2$ corresponds to the signal picked up by the microphone placed in the left ear. The short-time Fast Fourier Transforms (FFT) of the received noisy signals is formulated as follows:

$$Y_{i,N}(f, n) = S_{i,N}(f, n) + D_{i,N}(f, n) \quad i = \{1, 2\} \tag{2}$$

where $Y_{i,N}(f,n)$, $S_{i,N}(f,n)$ and $D_{i,N}(f,n)$ denote respectively the N-point FFTs of the $y_i(k)$, $s_i(k)$, and $d_i(k)$ for the frame n and the fth frequency bin. The parameter 'N' is left out for simplicity.

The proposed dual-channel speech enhancement algorithm contains then two major parts:

- Noise PSD estimation based on PSD and CPSD computation of received noisy speech signals.
- Enhanced speech signal estimation using spectral subtraction approach.

2.1. PSD and CPSD estimation

The noise PSD estimation needs a PSD and a CPSD estimation of the noisy received speech signals. The PSD of the noisy signal at the first channel '$P_{y1y1}(f,n)$', and at the second channel '$P_{y2y2}(f,n)$', and the CPSD '$P_{y1y2}(f,n)$' can be estimated as follows [19]:

$$\begin{aligned} P_{YiYi}(f, n) &= \lambda \cdot P_{YiYi}(f, n-1) + (1-\lambda) \cdot Y_i(f, n) \cdot Y_i^*(f, n) \quad i = \{1, 2\} \\ P_{Y1Y2}(f, n) &= \lambda \cdot P_{Y1Y2}(f, n-1) + (1-\lambda) \cdot Y_1(f, n) \cdot Y_2^*(f, n) \end{aligned} \tag{3}$$

where $*$ is the complex conjugate operator, λ is a smoothing factor usually close to 1. This factor should satisfy the two following constraints:

- For lower values, the estimation takes into account the speech short term stationarity.
- For higher values, this factor serves to minimize the estimator variance.

A previous study [9] showed that for 16 kHz sampling frequency with 256 samples per frame and a 50% overlap, the upper limit values of λ were around 0.6–0.8. In this study, we choose $\lambda = 0.8$.

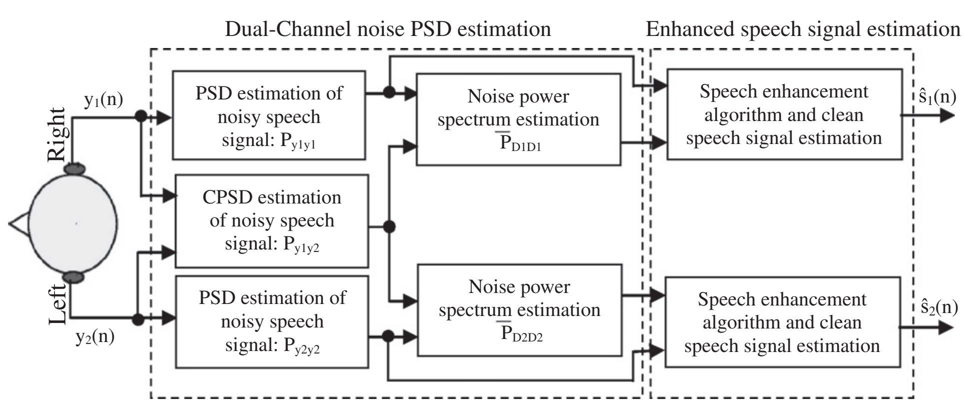

Fig. 1. Block diagram of the considered dual-channel speech enhancement algorithm.

2.2. Dual channel noise PSD estimation

For the derivation of the two-channel noise PSD estimator, the following assumptions are considered:

– The Speech and the noise signals are uncorrelated so that the CPSD between speech and noise signals $P_{SiDj}(f) = E\{S_i(f) \cdot D_j(f)^*\} = 0,\ \forall\ \{i,j\} = \{1,2\}^2$, where $E\{\cdot\}$ and $*$ denote respectively the expected value and the complex conjugate operators.
– The distance between both microphones is sufficiently high (180 mm) and the high coherence of the diffuse noise field at low frequencies is neglected. Therefore, the noise received by the microphones can be represented by two uncorrelated noise sources which implies that the noise CPSD $P_{D1D2}(f) = E\{D_1(f) \cdot D_2(f)^*\} = 0$.
– The speaker, placed in the frontal direction at 0 degree, is close to the two well calibrated omnidirectional microphones. The PSD of speech signals are then considered locally identical ($S1(f) = S_2(f) = S(f)$).

Based on the work of Dörbecker and Ernst [6], an extended dual channel noise PSD estimator is proposed tacking into account the difference between the PSD of the noise sources. The noise PSD in the right channel '$P_{D1D1}(f,n)$' and in the left channel '$P_{D2D2}(f,n)$' can be computed using the following equation:

$$\begin{aligned} P_{D1D1}(f,n) &= |P_{Y1Y1}(f,n)| - |P_{Y1Y2}(f,n)| \\ P_{D1D2}(f,n) &= |P_{Y2Y2}(f,n)| - |P_{Y1Y2}(f,n)| \end{aligned} \tag{4}$$

In order to improve the performance of the considered noise PSD estimator, a smoothing stage is considered. The estimated PSD of noise signals in the right and in the left channels are processed according to Eq. (5) in order to compute the smoothed noise PSDs ($\overline{P}_{D1D1}, \overline{P}_{D2D2}$).

$$\begin{aligned} \overline{P}_{D1D1}(f,n) &= \lambda' \cdot \overline{P}_{D1D1}(f,n-1) + (1-\lambda') \cdot P_{D1D1}(f,n) \\ \overline{P}_{D2D2}(f,n) &= \lambda' \cdot \overline{P}_{D2D2}(f,n-1) + (1-\lambda') \cdot P_{D2D2}(f,n) \end{aligned} \tag{5}$$

where λ' is a smoothing parameter in the range [0.5, 1]. In this study, λ' is fixed at 0.9.

The proposed noise PSD estimation is applied to the speech samples corrupted by babble noise at 0 dB. Fig. 2 depicts the real and the estimated noise PSDs for $f = 300$ Hz at the left and the right channels. It is interesting to observe the good noise PSD estimation on both channels.

2.3. Spectral subtraction algorithms

In our study, two spectral subtraction algorithms are considered and compared in the case of BCI. The first algorithm was described by Berouti et al. [1] where the noise power spectrum is multiplied by an over-subtraction factor α and subtracted from the noisy speech power spectrum in order to minimize the artefacts due to residual and musical noise. This algorithm is extended to a dual-channel non linear spectral subtraction (DC-NLSS) algorithm. The estimate of the clean speech power spectrum at the first and the second channel is given by the following equation:

$$|\widehat{S}_i(f)|^2 = \begin{cases} |Y_i(f)|^2 - \alpha_i \cdot \overline{P}_{DiDi}(f) & \text{if } |Y_i(f)|^2 > \beta \cdot \overline{P}_{DiDi}(f) \\ \beta \cdot \overline{P}_{DiDi}(f) & \text{else} \end{cases},\quad i = \{1,2\} \tag{6}$$

where 'i' is the microphone index, α_i ($\alpha_i > 1$) is the over-subtraction factor, which is a function of the segmental Signal-to-Noise Ratio (SSNR), and $\beta (0 < \beta < 1)$ is the spectral floor.

This implementation assumes that the noise affects the speech spectrum uniformly and the over-subtraction factor subtracts an overestimate of the noise over the whole spectrum. However, in real world noise, the noise spectrum is not uniform for all frequencies. For example, in the case of babble noise, most of the noise energy is concentrated in the low-frequency area. In order to take into account the fact that real world noise affects the speech spectrum differently at various frequencies, it becomes imperative to estimate a suitable factor that will subtract just the necessary amount of the noise spectrum from each frequency subband. That's why a multi-band spectral subtraction approach was proposed by Kamath and Loizou [15] and adopted by Udrea et al. [32]. In multi-band spectral subtraction approach, the noisy and the noise spectrum are divided into L non-overlapping bands, and spectral subtraction is performed independently in each frequency band. This multi-band spectral subtraction algorithm is also extended to a dual-channel configuration (DC-MBSS). Hence, the estimate

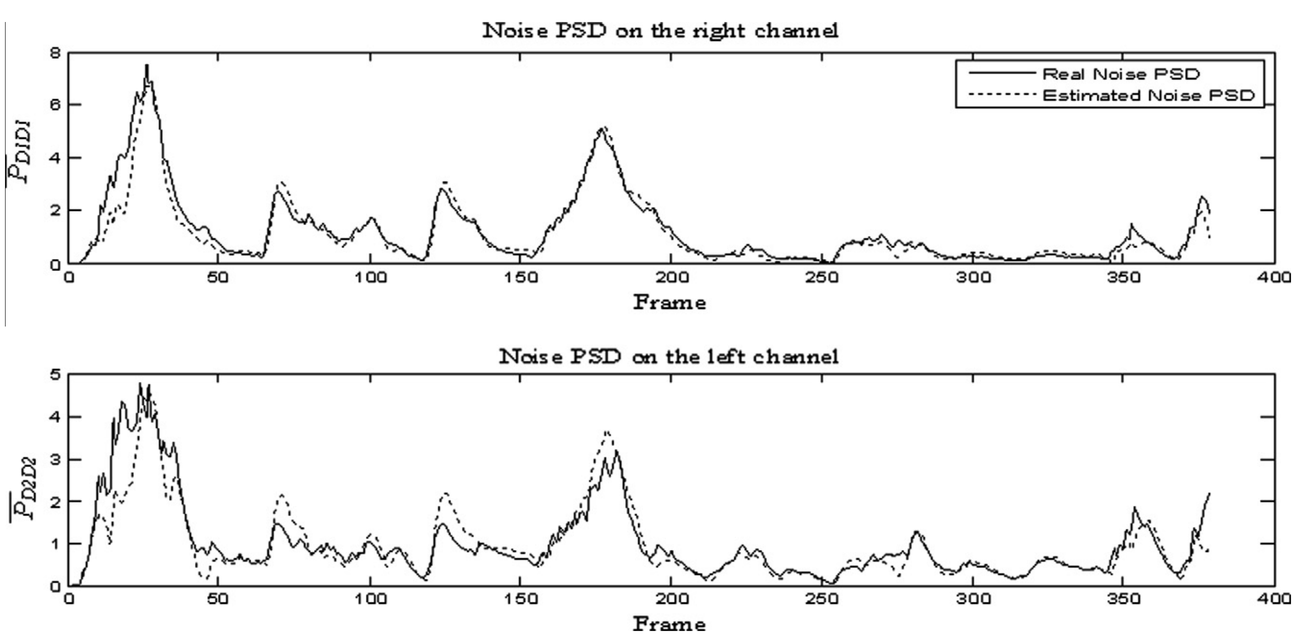

Fig. 2. Noise PSD and estimated noise PSD on the right (top) and the left (bottom) channels.

of the clean speech power spectrum in the lth band at the first and the second channels is given by the following equation:

$$|S_{il}(f)|^2 = |Y_{il}(f)|^2 - \alpha_{il} \cdot \delta_l \cdot \bar{P}_{l_{DiDi}}(f) \quad i = \{1,2\}, \quad b_l < f < e_l \tag{7}$$

where α_{il} is the over-subtraction factor of the lth frequency band at the ith channel and δ_l is a tweaking factor that can be individually set for each frequency band to customize the noise removal properties. b_l and e_l are the beginning and ending frequency bins of the lth frequency band.

The band specific over-subtraction factor α_{il} is a function of the SSNR_{il} of the lth frequency band which is calculated as given by the following equation:

$$\mathrm{SSNR}_{il}(\mathrm{dB}) = 10\log_{10}\left(\frac{\sum_{f=b_l}^{e_l} |Y_{il}(f)|^2}{\sum_{f=b_l}^{e_l} \bar{P}_{lDiDi}(f)}\right) \quad i = \{1,2\} \tag{8}$$

According to the SSNR_{il} values, the over-subtraction factor α_{il} is calculated as follows:

$$\alpha_{il} = \begin{cases} 4.75 & \mathrm{SSNR}_{il} < -5 \\ 4 - \frac{3}{20}\mathrm{SSNR}_{il} & -5 < \mathrm{SSNR}_{il} < 20 \quad i = \{1,2\} \\ 1 & \mathrm{SSNR}_{il} > 20 \end{cases} \tag{9}$$

While the use of the over-subtraction factor α_{il} provides a degree of control over the noise subtraction level for each frequency band, the use of multiple frequency bands and the use of the δ_l weights provide an additional degree of control within each frequency band. Since most of the speech energy is present in the lower frequencies, smaller δ_l values are used for low frequency bands in order to minimize speech distortion. The values of δ_l are empirically determined and set to following values [32]:

$$\delta_l = \begin{Bmatrix} 1 & 60\ \mathrm{Hz} < f_l < 300\ \mathrm{Hz} \\ 1.3 & 300\ \mathrm{Hz} < f_l < 1\ \mathrm{kHz} \\ 1.6 & 1\ \mathrm{kHz} < f_l < 2\ \mathrm{kHz} \\ 1.8 & 2\ \mathrm{kHz} < f_l < 3\ \mathrm{kHz} \\ 1.3 & 3\ \mathrm{kHz} < f_l < 8\ \mathrm{kHz} \end{Bmatrix} \tag{10}$$

The factors α_{il} and δ_l can be adjusted for each frequency band and for different speech conditions to get better speech quality. Finally, the negative values in the modified power spectrum calculated according to Eq. (7) are floored to the noisy power spectrum:

$$|\hat{S}_{il}(f)|^2 = \begin{cases} |\hat{S}_{il}(f)|^2 & if\,|\hat{S}_{il}(f)|^2 > \beta \cdot \bar{P}_{lDiDi}(f) \\ \beta \cdot P_{lDiDi}(f) & \text{otherwise} \end{cases}, \qquad i = \{1,2\}, \quad b_l < f < e_l \tag{11}$$

where the spectral floor parameter is set to $\beta = 0.01$. The modified spectra of each frequency band $(|\hat{s}_{il}(f)|)$ are recombined to obtain the enhanced speech spectrum $(|\hat{s}_i(f)|)$. This modified spectrum is combined with the phase information of the noisy input signal $(\theta_{Yi}(f) = angle(Y_i(f)))$ to reconstruct the time speech signal by using the Inverse Fast Fourier Transform (IFFT) in conjunction with the overlap and add method. The enhanced output signals on both channels are then computed according to the following equation:

$$\hat{s}_i(k) = \mathrm{IFFT}(|\hat{S}_i(f)| \cdot e^{j\theta_{Yi}(f)}) \quad i = \{1,2\} \tag{12}$$

3. Bilateral cochlear implant simulator

In order to investigate the behavior of the above SEAs in the case of cochlear implant, vocoder stimulations were used. However, it was shown by many (e.g. [35]) that these simulations provide results consistent with the outcome of cochlear implants and that vocoded speech signals could be presented to normal hearing listeners in the absence of confounding factors associated with cochlear implants.

Fig. 3 gives the block diagram of the considered CI simulator. The conception of this simulator is based on the 12-channel Neurelec Digisonic SP cochlear implant. Both left and right enhanced speech signals are processed with the same CI simulator.

The enhanced speech signal, sampled at 16 kHz, is divided into 50% overlapping frames by the application of a Hanning window function. The length of the window (frame) is set to $V = 128$ samples. FFT is performed on the previously windowed speech signal (input blocks of $V = 128$ samples). The obtained short time spectrum is passed through a bank of band-pass filters.

According to psychoacoustic evidence, cochlear tonotopy follows a lin-log scale (linear in low frequencies and logarithmic in high frequencies) which can be simulated by a bark scale. So, the cut-off frequencies between the different frequency bands are chosen in order to respect this scale. The mapping between the bark scale and the frequencies scale is non-linear according to the non-linearity of the human ear. An approximate analytical expression for describing the conversion from linear frequency f, into the critical band number B in bark is given by the following equation [31]:

$$B(f) \approx 6.7 * \mathrm{arsinh}\left(\frac{f - 20}{600}\right) \tag{13}$$

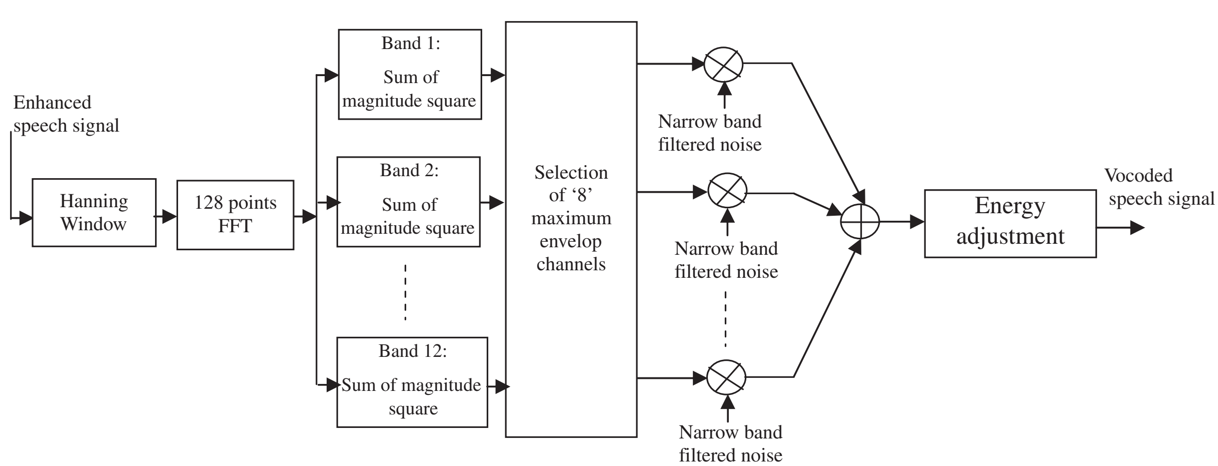

Fig. 3. Block diagram of the adopted CI simulator.

Table 1
Number of FFT bins related to each frequency analysis band and its associated center frequencies in Hz.

Analysis frequency band 'z'	1	2	3	4	5	6	7	8	9	10	11	12
Left and right channel												
Cut-off frequencies (Hz)	300–448	448–616	616–810	810–1041	1041–1318	1318–1654	1654–2065	2065–2568	2568–3187	3187–3950	3950–4892	4892–6055
Number of lines N_z	1	1	2	2	2	3	3	4	5	6	8	9
Starting line $nstart_z$	3	4	5	7	9	11	14	17	21	26	32	40
Center freqs f_{mid} (Hz)	375	500	687	937	1187	1500	1875	2312	2875	3562	4437	5500

The bandwidth of those filters matched the frequency range corresponding to the cochlear segment excited by electrical stimulation. Table 1 presents the cut-off frequencies, the number of lines (N_z), the starting line ($nstart_z$) assigned to each frequency band 'z' ($z = 1, \ldots, 12$) and its associated center frequency (f_{mid}).

The power in the frequency analysis band 'z' ($E(z)$) can be computed using the following power estimation equation:

$$E(z) = \frac{\sum_{f=nstart_z}^{f=nstart_z+N_z-1} e(f)}{N_z}, \quad z = 1, \ldots, 12 \tag{14}$$

where $e(f)$ is the power levels of the frequency bin 'f'. The output of this analysis is a vector of power values for each frame of data. According to the advance combination encoder (ACE) strategy [4] and for each frame, only the first eight frequency bands presenting the most important power levels are used for vocoded speech signal reconstruction. The power levels of the remaining bands are set to zero.

For each frequency analysis band 'z', a Hanning window '$w(v)$' is weighted by its related power value '$E(z)$' to get a modified Hanning window called 'the envelope: $Env(z,v)$' according to the following equation:

$$Env(z, v) = E(z) \cdot w(v), \quad z = 1, \ldots, 12, \quad v = 1, \ldots, V \tag{15}$$

To prevent sharp variations, a smoothing is performed throughout the bands using a further low pass filtering with a 150 Hz cutoff frequency.

In order to get the carrier synthesis noise signals, a white noise is shaped to each frequency analysis band by a 3rd order band-pass Butterworth filter. For each frequency band, a vocoded speech signal is synthesized by modulating the obtained narrow band white noise carrier with its corresponding smoothed envelope. Finally, all the synthesized vocoded signals coming from the different frequency analysis bands are summed up and the output is adjusted to have the same long-term root mean square energy as the input speech signal (70 dB).

4. Performance evaluation

In this section, the proposed speech enhancement algorithms are evaluated using both objective measures and subjective listening tests.

4.1. Phonetic material

The used phonetic material is the French Lafon set [21] which contains twenty lists each one composed of seventeen 3-phoneme words pronounced by a single male talker. This is a commonly used speech stimuli for intelligibility assessment in French. Sound level was calibrated to 70 dB SPL. All theses lists are processed in the anechoic room of the ORL department of the Edouard-Herriot Hospital of Lyon-France with an additive babble noise. The SNRs varied from −3 dB to 6 dB with 3 dB steps.

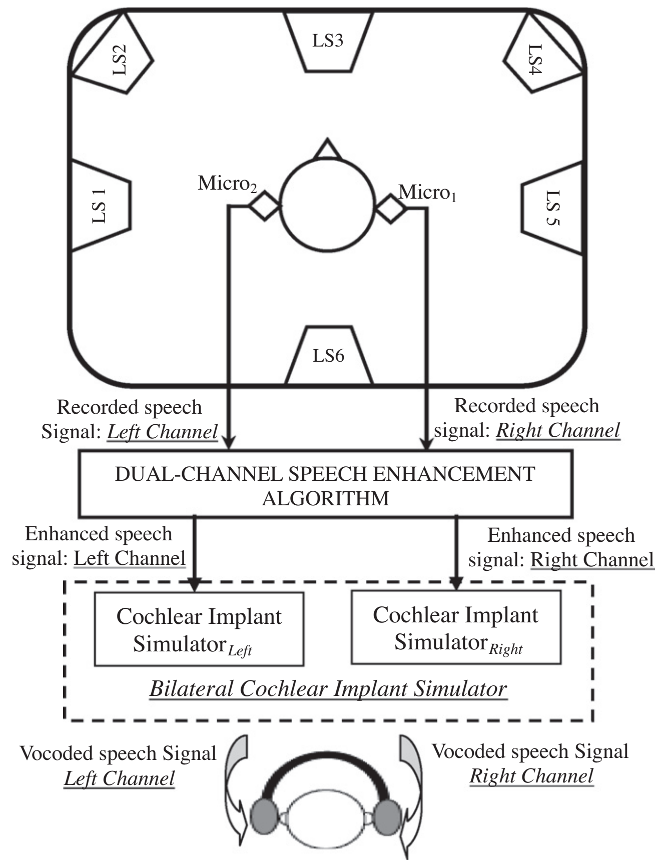

Fig. 4. Experimental setup (Anechoic room) LS_3 = speech signal, LS_2, LS_4 and LS_5 = noise signal, LS_1 and LS_6 were not used.

The experimental setup is presented in Fig. 4. The target speech signal is always placed directly in front of the listener (a KEMAR mannequin) at 0° azimuth (LS_3 position). Speech signals are corrupted with either one or three noise interferers. In the single interferer condition, a single speech babble noise source is presented from the right side of the listener (LS_5 position). When multiple interferers are present, three interfering noise sources are placed asymmetrically either across both hemifields (−60°, 60°, and 90° corresponding respectively to LS_2, LS_4 and LS_5 positions). The speech stimuli are processed offline with MATLAB software.

4.2. Objective measures

The performance of the previously described DC-MBSS algorithm is evaluated with different number of bands varying from one to eight (L = 1–8). The speech spectrum is split into different bands mapped to the bark scale as described is Section 3. The number of FFT bins (N_L) and the start bin ($nstart_L$) of each frequency

Table 2
Number of FFT bins related to each band for different number of frequency analysis bands.

Number of analysis frequency bands 'L'		Band 1	2	3	4	5	6	7	8
2	N_L	11	53						
	$nstart_L$	1	12						
3	N_L	6	15	43					
	$nstart_L$	1	7	22					
4	N_L	4	7	16	37				
	$nstart_L$	1	5	12	28				
5	N_L	3	5	8	17	31			
	$nstart_L$	1	4	9	17	34			
6	N_L	2	4	5	10	15	28		
	$nstart_L$	1	3	7	12	22	37		
7	N_L	2	3	4	6	9	15	25	
	$nstart_L$	1	3	6	10	16	25	40	
8	N_L	2	2	3	4	7	9	15	22
	$nstart_L$	1	3	5	8	12	19	28	43

band are recapitulated in Table 2 for different number of frequency analysis bands.

The two following objective measurements are considered for performance assessment:

– Perceptual Evaluation of Speech Quality (PESQ) score which ranges from 0.5 (for the worst case) to 4.5 (for the best case) according to the ITU-T Recommendation P. 862 standard [12].
– Itakura-Saito (IS) distance which is based on the similarity or difference between the all-pole model of the clean and the enhanced speech signals [24].

This experiment is performed with only one interfering babble noise source at 0 dB SNR. Results are shown in Fig. 5a and Fig. 5b respectively indicating the mean PESQ score and IS distance versus the number of frequency bands for both left and right channels.

From Fig. 5a, we observe that the PESQ scores increased significantly with the number of frequency bands up to 6 for both channels. Further increase in the number of bands is not found to be beneficial. Additionally, we can see that the PESQ scores for the left channel (PESQ-left) are greater than the PESQ scores for the right channel (PESQ-right). This can be explained by the position of the interfering noise source which is closer to the right ear. On the other hand, the IS values, given by Fig. 5b, decreased with the number of bands for both channels. This is because most of the residual noise is reduced leading to an improvement in speech quality.

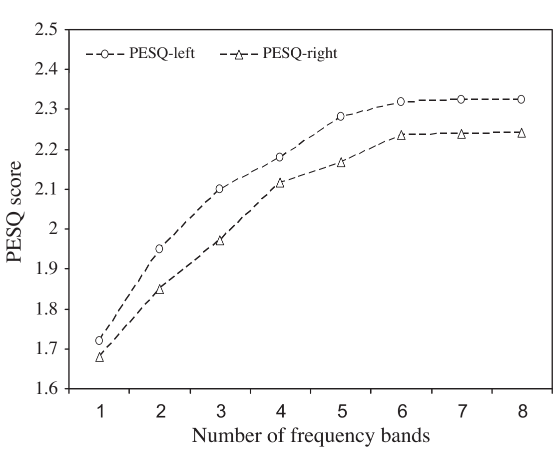

Fig. 5a. Mean PESQ scores in right and left channel.

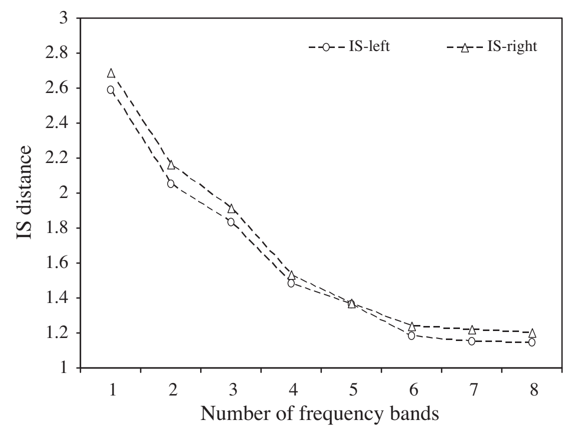

Fig. 5b. Mean IS values in right and left channel.

4.3. Subjective evaluation

In this section, we investigate the potential benefits of processing the noisy speech signal with the proposed DC-NLSS and DC-MBSS algorithms using six frequency analysis bands. Therefore, to assess the performance of the aforementioned noise reduction algorithms, comprehensive phoneme recognition tests are firstly conducted in BCI simulation with normal hearing listeners and secondly with BCI implantees.

4.3.1. Listeners

The performance of the proposed speech enhancement algorithms is evaluated with a population of 50 normal hearing subjects and 3 postlingually implantees. Normal hearing subjects ranged from 18 to 32 years. BCI subjects are fitted bilaterally with the binaural Digisonic SP multichannel implant device manufactured by Neurelec Corporation (France) and their biographical data are indicated in Table 3. All participants were native French speaking subjects. Listening sessions took place in the Cochlear Implant Room of the Edouard-Herriot Hospital. Subjects' hearing is checked prior to the experiment.

Table 3
Biographical data of recruited BCI implantees.

	Subject 1	Subject 2	Subject 3
Age (years)	56	36	48
Gender	F	M	F
Past surgery (years)	4	4	4

4.3.2. Listening session procedure

The listening tests are conducted using a personal computer connected to a CD player (PHILIPS-CD723). An audiometer (MADSEN-Orbiter 922) is used for calibration and intensity level adjustment. Stimuli are presented bilaterally to the subjects using a closed professional 'Sennheiser' HD250 linear headphones at a comfortable level calibrated to 70 dB SPL.

Prior to the formal testing, a training session containing 10 random words is administrated in order to familiarize each subject with the stimuli. No scores are calculated in the training session. Following the practice session, the subjects were tested in the various experimental conditions. During the testing session, the subjects are instructed to repeat the words they heard. In total, there are 24 testing conditions (3 processing methods $\times$ 4 SNRs $\times$ 2 noise interfering configurations). A list of 17 words is used for each of the 20 firstly presented conditions. Four randomly selected lists are reutilized for the remaining conditions. A sequential test order, starting with lists processed in noise from the highest SNR (6 dB) to the lowest SNR (-3 dB) is employed. We take this sequential approach in order to give the subjects some time to adapt to the listening in noisy conditions. At the end of each listening session, the responses of each subject are collected, stored and scored off-line with the percentage of correctly repeated phonemes.

4.3.3. Results

The performance of the proposed speech enhancement algorithms is carried out through listening tests. Tests are conducted in BCI simulation with normal hearing listeners and BCI implanted subjects. Intelligibility scores for this experiment are derived from the percentage of correctly repeated phonemes per test condition. Chi2 tests with the following parameters are performed to examine the effects of the considered factors:

– Repeated measures (the same subjects underwent all the situations)
– Dependant variable (Recognition score in percent)
– Three factors:
  • Method (Noisy, DC-NLSS, DC-MBSS)
  • Noise interfering configuration ('90°' and '−60°, 60°, 90°')
  • SNR (-3 to 6 dB with 3 dB step, this last factor was taken as random; the first two factors were fixed).

This model, with both fixed and random effects, is known as mixed-effect model, and is a standard for data analysis [2]. We fit our models using the 'lmer' program of the 'lme4' package [3] of the statistical programming language and software environment R [27].

4.3.3.1. Results with normal hearing subjects. Fig. 6 shows the experimental results obtained with normal hearing subjects. Results revealed a significant main effect of the method (Chi2[2] = 526, $p < 0.001$), of the SNR (Chi2[3] = 4178, $p < 0.001$) and of the noise interfering configuration (Chi2[1] = 97, $p < 0.001$). In addition, a significant interaction is observed between the method and the SNR (Chi2[6] = 45, $p < 0.001$). No significant interaction between the method and the noise interfering configuration (Chi2[2] = 3, $p = 0.122$) and between noise interfering configuration and the SNR (Chi2[3] = 5, $p = 0.4$) are noticed.

Post-hoc comparisons are run to assess the significant improvements in scores obtained with the considered methods at different SNRs. We use the Tukey HSD test via the general linear hypothesis test '*glht*' function from multiple comparison '*multcomp*' package of R. The mean recognition scores obtained using the DC-NLSS speech enhancement method are significantly higher ($p < 0.001$) than the scores obtained with the unprocessed noisy speech signals for all SNRs. Furthermore, the mean recognition scores achieved with the DC-MBSS method are significantly higher than those obtained with the DC-NLSS method ($p < 0.001$) at 3, 0, and -3 dB SNRs; but do not differ significantly at 6 dB ($p = 0.142$).

4.3.3.2. Results with BCI implantees. Experimental results obtained with three BCI implantees are presented in this section. Fig. 7a and Fig. 7b show the individual subject scores obtained with three methods when single and three interfering noise sources are respectively considered.

Statistical tests indicate a main significant effect of the method (Chi2[2] = 70, $p < 0.001$), of the SNR (Chi2[3] = 206, $p < 0.001$) and of the noise interfering configuration (Chi2[1] = 10, $p < 0.001$). However, there is no significant interaction between the method and the SNR (Chi2[6] = 3, $p = 0.7$) and between the method and the noise interfering configuration (Chi2[2] = 1, $p = 0.8$).

5. Discussion

In the present study, we assessed the benefit of two proposed DC-NLSS and DC-MBSS SEAs. A subjective comparative study was considered by computing the mean recognition scores at different SNRs and noise interfering configurations. Performance evaluation

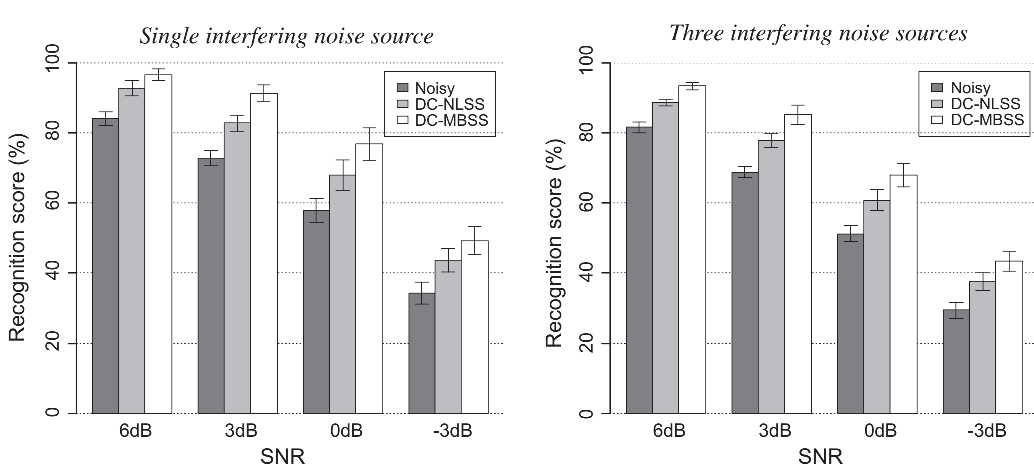

Fig. 6. Recognition score (mean ± standard error) in the presence of babble noise with noisy speech, DC-NLSS and DC-MBSS enhancement algorithms for two noise interfering configurations.

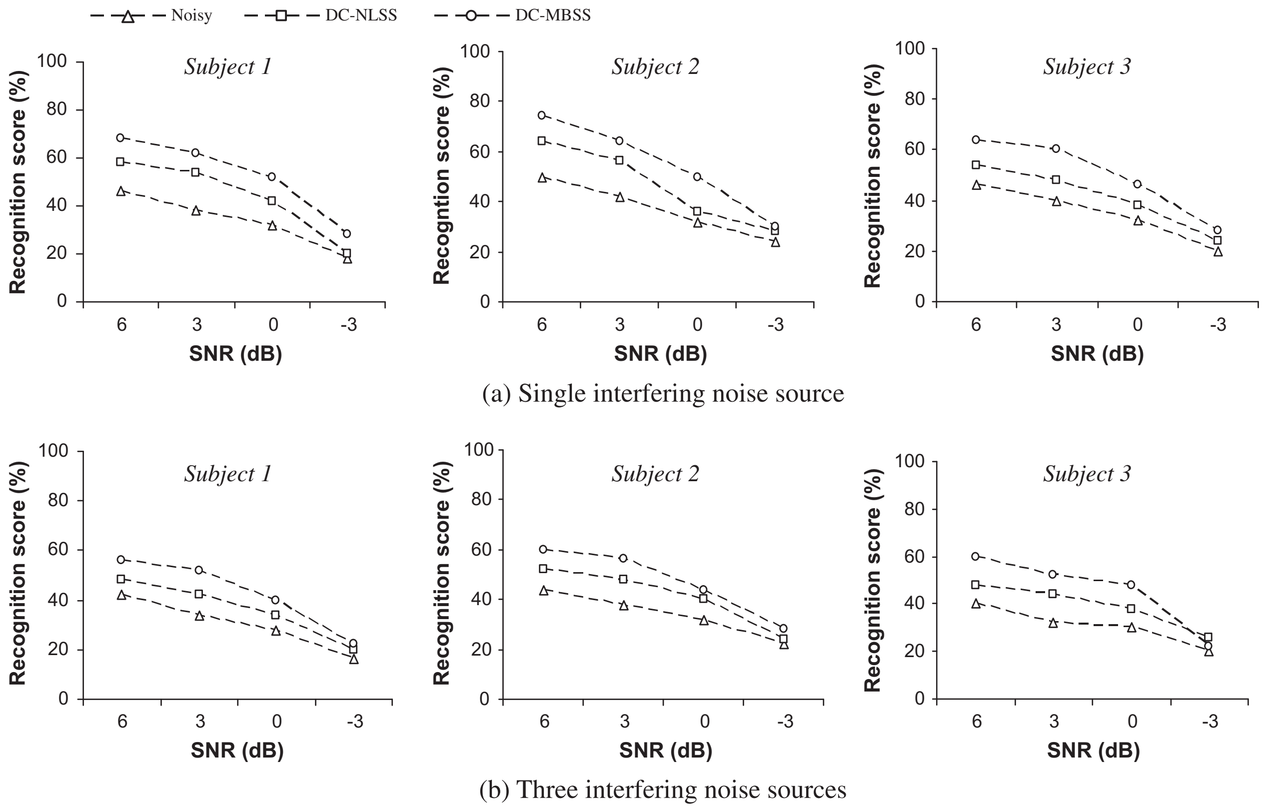

Fig. 7. Phoneme's recognition scores with deafened BCI subjects for different methods at all SNRs.

was conducted with normal hearing listeners using a specific BCI simulator and with BCI implantees.

5.1. Speech enhancement algorithms vs SNRs

As shown in Fig. 6, compared with recognition score obtained using unprocessed noisy speech, an average improvement of 13% and 7% was achieved respectively with DC-MBSS and DC-NLSS algorithms. It was worthwhile noticed that this improvement was fairly constant at 6 dB and 3 dB, but was decreased for lower values of SNR (0 dB and −3 dB). This behavior was confirmed with the BCI implantees (Fig. 7) where an average improvement in recognition score between 9% (at positive SNR) and 4% (at −3 dB SNR) was observed when DC-NLSS was considered. It is also interesting to note the reasonably better performance of the DC-MBSS over DC-NLSS where the average improvement was between 8% (at positive SNR) and 6% (at −3 dB SNR). To sum up, our results revealed the outperformance of the DC-MBSS over the DC-NLSS especially for high SNRs. This may be due to the fact that DC-MBSS takes into account that noise signal affects the speech spectrum differently at various frequencies, and the over-subtraction coefficients are calculated at each frequency band.

5.2. Speech enhancement algorithms vs interfering noise source

First, the performance of speech enhancement algorithms was evaluated with normal hearing subjects using a BCI simulator. In the case of a single interfering noise source, the average improvement was about 8% and 16% for respectively DC-NLSS and DC-MBSS speech enhancement algorithms and this when compared to the scores obtained with the vocoded noisy speech signals. The Average improvement dropped to 6% and 13% for respectively DC-NLSS and DC-MBSS SEAs in the presence of three interfering noise sources.

Next, the same experiment was performed with three BCI implantees. When the DC-MBSS algorithm was applied, we observed a small benefit in term of recognition score from 14% when three interfering noise sources were present to around 18% in the presence of only one interfering noise source. This improvement in performance was a bit smaller for the case of the DC-NLSS algorithm.

As reported, the improvement in recognition score was smaller when multiple noise sources were present. This could be in part attributed to the fact that, in the binaural hearing literature, when both CI devices are available, BCI listeners can benefit from the head-shadow effect which may occur mainly when speech and noise are spatially separated [30,10,20].

6. Conclusion

Two speech enhancement algorithms based on spectral subtraction approach were proposed and implemented. The enhanced speech signals were obtained using either Dual Channel Non Linear Spectral Subtraction 'DC-NLSS' or Dual-Channel Multi-Band Spectral Subtraction 'DC-MBSS' algorithms which were successfully applied to the bilateral cochlear implant case.

Several phoneme recognition experiments were performed using the French Lafon set made of 20 lists. Each list was composed of seventeen 3-phoneme words pronounced by a single male talker. Speech signals were corrupted by an additive babble noise at different SNR levels and in the presence of single and multiple noise interfering sources. A total of 50 normal hearing subjects and 3 BCI implantees participated in the experiments.

Objective measures based on Perceptual Evaluation of Speech Quality (PESQ) score and Itakura-Saito (IS) distance indicated that the performance of the DC-MBSS algorithm did not improve once the number of bands exceeds 6. Then, subjective preference tests revealed that dual-channel speech enhancement algorithms and particularly DC-MBSS could yield substantial benefits in speech intelligibility for BCI users, especially in the presence of only one interfering noise source.

Acknowledgments

This research work is supported by the Franco-Tunisian cooperation project 'PHC-Utique' managed by CMCU (2008-2011). We appreciate the efforts of all the subjects. We also would like to thank the Cochlear Implantation Center in the Edouard Herriot hospital in Lyon France for the technical support and for the assistance in the listening test process.

References

[1] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. In: Proc IEEE int conf acoust, speech, signal process; 1979. p. 208–11.
[2] Baayen RH. Analyzing linguistic data. In: A practical introduction to statistics. Cambridge, UK: Cambridge University Press; 2008.
[3] Bates DM. Linear mixed model implementation in lme4. Manuscript, University of Wisconsin-Madison; 2007.
[4] Cochlear Corporation. ACE and CIS DSP strategies. Software requirements specification. Lane Cove, New South Wales, Australia; 2002.
[5] Chung K, Zeng FG, Acker KN. Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. J Acoust Soc Am 2006;120:2216–27.
[6] Dörbecker M, Ernst S. Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverbcration. In: European signal processing conference (EUSIPCO-96); 1996.
[7] Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble W. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. Ear Hear 2008;29:352–9.
[8] Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble1 W. Bilateral and unilateral cochlear implant users compared on speech perception in noise. Ear Hear 2010;31:296–8.
[9] Guérin A, Le Bouquin-Jeannès R, Gérard F. A two-sensor noise reduction system: applications for hands-free car kit. EURASIP J Appl Signal Process 2003;11:1125–34.
[10] Hawley ML, Litovsky RY. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J Acoust Soc Am 2004;115:833–43.
[11] Hu Y, Loizou P, Li N, Kasturi K. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. J Acoust Soc Am 2007;122:128–34.
[12] ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs; 2001.
[13] Kompis M, Dillier N. Noise reduction for hearing aids: combining directional microphones with an adaptive beamformer. J Acoust Soc Am 1994;96:1910–3.
[14] Kasturi K, Loizou P. Use of s-shaped input-output functions for noise suppression in cochlear implants. Ear Hear 2007;28:402–11.
[15] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In: Proceedings of ICASSP-2002, Orlando; 2002.
[16] Kokkinakis K, Loizou PC. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. J Acoust Soc Am 2010;127:3136–44.
[17] Loizou P, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. J Acoust Soc Am 2005;118:2791–3.
[18] Loizou PC, Hu Y, Litovsky P, Yu G, Peters R, Lake J, et al. Speech recognition by bilateral cochlear implant users in a cocktail party setting. J Acoust Soc Am 2009;125:372–83.
[19] Le Bouquin-Jeannès R, Faucon G. Using the coherence function for noise reduction. IEEProc 1992;139:276–80.
[20] Litovsky RY, Parkinson A, Arcaroli J. Spatial hearing and speech intelligibility in bilateral cochlear implant users. Ear Hear 2009;30:419–31.
[21] Lafon JC. Le test phonétique et la mesure de l'audition. In: Eindhoven (Eds.), Centrex; 1964.
[22] Monica LH, Ruth YL. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J Acoust Soc Am 2004;115:833–43.
[23] Noble W, Tyler R, Dunn CC, Bhullar N. Hearing handicap ratings among different profiles of adult CI users. Ear Hear 2008;29:112–20.
[24] Quackenbush S, Barnwell T, Clements M. Objective Measures for speech quality testing. Prentice-Hall; 1988.
[25] Qin MK, Oxenham AJ. Effects of simulated cochlear implant processing on speech reception in fluctuating maskers. J Acoust Soc Am 2003;114:446–54.
[26] Ricketts TA, Grantham DW, Ashmead HD, Haynes DH, Labadie RF. Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources. Ear Hear 2006;27:763–73.
[27] R development core team. R: A language and environment for statistical computing. Vienna: R; 2007.
[28] Spahr A, Dorman M. Performance of patients fit with advanced bionics CII and nucleus 3G cochlear implant devices. Arch Otolaryngol Head Neck Surg 2004;130:624–8.
[29] Spriet A, Van Deun L, Eftaxiadis K, Laneau J, Moonen M, Van Dijk B, et al. Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the nucleus freedom cochlear implant system. Ear Hear 2007;28:62–72.
[30] Tyler RS, Dunn CC, Witt S, Preece JP. Update on bilateral cochlear implantation. Curr Opin Otolaryngol Head Neck Surg 2003;11:388–93.
[31] Traunmüller H. Analytical expression for the tonotopic sensory scale. J Acoust Soc Am 1990;85:97–100.
[32] Udrea RM, Ciochina SD, Vizireanu N. Multi-band Bark scale spectral over-subtraction for colored noise reduction. In: Proceedings of "ISSCS," Ias-i, Romani, vol. 1. 2005. p. 311–4.
[33] Van Hoesel RJ. Exploring the benefits of bilateral cochlear implants. Audiol Neurootol 2004;9:234–46.
[34] Van Hoesel R, Clark G. Evaluation of a portable two microphone adaptive beamforming speech processor with cochlear implant patients. J Acoust Soc Am 1995;97:2498–503.
[35] Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. J Acoust Soc Am 2007;122:2376–88.
[36] Yang LP, Fu QJ. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. J Acoust Soc Am 2005;117:1001–4.

Applied Acoustics 73 (2012) 256–264

Contents lists available at SciVerse ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A noise cross PSD estimator based on improved minimum statistics method for two-microphone speech enhancement dedicated to a bilateral cochlear implant

Fathi Kallel [a,b,*], Mohamed Ghorbel [a], Mondher Frikha [a], Christian Berger-Vachon [b], Ahmed Ben Hamida [a]

[a] Research Unit in Advanced Technologies for Medical and Signals (ATMS), Laboratory of Electronics and Information Technologies (LETI), National Engineering School of Sfax, University of Sfax, Route Soukra km 3, B.P.W, 3038 Sfax, Tunisia

[b] PACS Team, INSERM Unit 1028: "Cognition and Brain Dynamics", Lyon Neurosciences Center, EPU-ISTIL, Claude Bernard University, Boulevard du 11 Novembre 1918, Villeurbanne 69622, France

ARTICLE INFO

Article history:
Received 10 December 2010
Received in revised form 18 September 2011
Accepted 20 September 2011
Available online 19 October 2011

Keywords:
Dual-microphone noise reduction
Noise CPSD estimation
Minima tracking
BCI

ABSTRACT

Coherence based methods have been successfully applied to dual-microphone noise reduction systems. These techniques showed good results when noise signals on two microphones were uncorrelated, but their performance decreased with correlated noises. It could be improved when the cross power spectral density (CPSD) of received noises is available.

In this paper, an improved minimum tracking (IMT) technique for noise CPSD estimation was proposed. The performance of this technique was compared to two other noise CPSD estimators based on voice activity detection (VAD) and minimum tracking (MT) approaches. Evaluation was performed at four signal-to-noise ratios (SNR) and two interfering noise source configurations.

Results showed a superiority of the IMT approach in terms of low computing time and quality indicated by the perceptual evaluation of speech quality (PESQ) scores. Then, subjective listening tests were carried out with 50 normal hearing listeners using a specific bilateral cochlear implant (BCI) simulator and utilizing the French Lafon database corrupted by additional babble noise. Results obtained with the proposed technique were better than the two previously mentioned noise CPSD estimators.

1. Introduction

In quiet listening conditions most cochlear implant (CI) users can now achieve even more than 80% word recognition scores regardless the used device [3]. However, speech recognition scores are highly degraded in noisy environments [25]. Furthermore, as mentioned by CI users, better and comfortable speech recognition in noisy environments would be considered as one of the most significant challenges [43]. Now, to improve speech intelligibility in noisy environments, individuals with severe to profound hearing loss could be implanted with two cochlear implants, one in each ear. In fact, bilateral cochlear implantation provides patients the advantages of bilateral information. Bilateral hearing permits a better understanding of speech in quiet and also in noisy environments [5,20]. However, under more challenging listening conditions, BCI recipients perform poorly, compared to normal-hearing listeners [31].

Numerous speech enhancement algorithms, divided into single-microphone and multi-microphone methods, have been proposed over the years to improve speech recognition in noisy background conditions. Different single-microphone algorithms, originally developed for normal hearing listeners, have been applied to CI speech processing [16,19,32,45]. These algorithms were able to improve CI users' performance in noisy conditions, but they have limitations in real environments. They introduce musical noise and speech distortion [13]. Considerably larger benefits in speech intelligibility could be obtained when resorting to multi-microphone adaptive signal processing strategies, instead. Such strategies make use of spatial information due to the relative position of the emanating sounds; and they exploit situations in which target and masker are spatially separated [15,21,39]. However, a larger number of microphones implies higher costs and an increase computational load. Dual-microphone approaches give a trade off between multi and single microphone methods. They show some interest in different promising applications like hands-free systems [2,28], hearing aids [17,29] and cochlear implant [8, 14,].

Kallel et al. [8] proposed two speech enhancement algorithms based on spectral subtraction principle for bilateral cochlear implant users, specifically, dual-channel non linear spectral subtraction and dual-channel multi-band spectral subtraction algorithms. Subjective preference tests revealed that dual-channel

* Corresponding author at: Research Unit in Advanced Technologies for Medical and Signals (ATMS), Laboratory of Electronics and Information Technologies (LETI), National Engineering School of Sfax, University of Sfax, Route Soukra km 3, B.P.W, 3038 Sfax, Tunisia.

E-mail address: fathikallel@yahoo.fr (F. Kallel).

0003-682X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2011.09.008

speech enhancement algorithms and particularly dual-channel multi-band spectral subtraction algorithm could yield substantial benefits in speech intelligibility for BCI users. Performance of adaptive beamformer techniques with multi-microphone for BCI users was also assessed by different studies [18,26,42].

Coherence-based methods are known as a subclass of dual-microphone methods. They provide good results in uncorrelated noise environments. The drawback of these methods is a decrease in performance if captured noises are correlated. To handle this problem, Le Bouquin-Jeannès et al. [34] proposed computing the speech CPSD by subtracting the noise CPSD from the noisy speech CPSD. The main difficulty of this improvement is to estimate the CPSD of received noise signals.

Guérin et al. [2] proposed a noise CPSD estimation method as a function of the posteriori SNR and the CPSD of the noisy signal based on the premise that the noise CPSD can be estimated in all frames. Zhang and Jia [44] proposed a decision based method for noise CPSD estimation during speech pauses. They employed minimum statistics approach on each channel to estimate the noise power spectral density (PSD) in that channel. They exploited the estimated noise PSDs as a criterion to distinguish speech and pause frames. Rahmani et al. [23] proposed another method to estimate the noise CPSD without using a VAD or other conventional noise estimation techniques. This method is based on the phase information of the noisy speech signals. Different single-microphone algorithms originally developed for noise PSD estimation were extended to a dual-channel case for noise CPSD estimation. Sovka et al. [33] proposed a single-microphone algorithm for noise PSD estimation based on an iterative method. Rahmani et al. [22] proposed an extension of this algorithm to a dual-microphone case for noise CPSD estimation which is updated in all speech and non-speech frames. In this iterative method, the noise CPSD is estimated using the gain filter of the previous frame. Rahmani et al. [24] proposed also an extension of a method introduced by Martin [37,38] to a dual-microphone approach to estimate the noise CPSD. In this method, the noise CPSD is estimated by tracking the minimum of the noisy speech over a search window spanning a specific number of frames.

The present study proposes a noise CPSD estimation method using minimum tracking (MT) technique. In this method, a new non-linear rule for tracking the minimum of the noisy speech CPSD is applied by continuously averaging past spectral values. This approach differs from the method proposed by Rahmani et al. [24] in which the noise CPSD estimation was dependent on the length of the minimum search window. This proposed method is an extension of the single-microphone approach previously proposed by Farsi [9] for noise PSD estimation to the dual-microphone case.

The current paper is outlined as follows. Section 2 provides theoretical overview of two-microphone speech enhancement system and both VAD and MT based methods for noise CPSD estimation. The proposed new noise CPSD estimation method is presented in Section 3. Section 4 derives bilateral cochlear implant simulator principle. Section 5 evaluates the experimental results and gives an overall discussion of all obtained results. Section 6 concludes the paper.

2. Dual-microphone speech enhancement system

2.1. General consideration

The bilateral cochlear implant configuration is illustrated in Fig. 1. Both left and right cochlear implants are fitted with one microphone. The two received noisy speech signals ($y_1(n)$ and $y_2(n)$) are processed together with the proposed speech enhancement algorithm (SEA) to generate enhanced speech signals ($\hat{s}_1(n)$ and $\hat{s}_2(n)$) which are then used in stimulation. The proposed dual-channel SEA contains two major parts: Noise CPSD estimation and enhanced speech signal computing using the cross power spectral subtraction (CPSS) algorithm.

The picked up noisy speech signals could be expressed in time domain as follows:

$$y_i(k) = s_i(k) + d_i(k) \quad i = \{1,2\} \tag{1}$$

where 'i' is the microphone index, $y_i(k)$, $s_i(k)$ and $d_i(k)$ represent respectively noisy speech, clean speech and noise signals and k is the discrete time. Note that $i = 1$ corresponds to the signal picked up by the microphone placed in the right ear and $i = 2$ corresponds to the signal picked up by the microphone placed in the left ear.

The short-time discrete Fourier transform (STDFT) of the received noisy signals is formulated as follows:

$$Y_{i,N}(f,n) = S_{i,N}(f,n) + D_{i,N}(f,n) \quad i = \{1,2\} \tag{2}$$

where $Y_{i,N}(f, n)$, $S_{i,N}(f, n)$ and $D_{i,N}(f, n)$ represent respectively noisy speech, clean speech and noise signals in the frequency domain, N is the frame length, $f \in [0, N-1]$ is a discrete variable enumerating the frequency lines, and n is a time frame number. The parameter 'N' is left out for simplicity.

The coherence between the noisy speech signals $y_1(k)$ and $y_2(k)$, noted $\Gamma_{Y1Y2}(f, n)$, is computed according to Eq. (3) [35] and is situated between 0 and 1. A spectral modification filter (H_{cpss}) can be computed as a function of this coherence function.

$$\Gamma_{Y1Y2}(f,n) = \frac{|P_{Y1Y2}(f,n)|}{\sqrt{P_{Y1Y1}(f,n) \cdot P_{Y2Y2}(f,n)}} \tag{3}$$

where $P_{Y1Y1}(f, n)$, $P_{Y2Y2}(f, n)$ and $P_{Y1Y2}(f, n)$ are the PSD of $Y_1(f, n)$ and $Y_2(f, n)$, and CPSD between $Y_1(f, n)$ and $Y_2(f, n)$, respectively.

As can be seen in Fig. 1, the computed filter $H_{cpss}(f, n)$ is applied on both noisy signals to obtain enhanced signals using an inverse short-time discrete Fourier transform (ISTDFT).

The estimated CPSD and PSDs of noisy speech signals are computed using a first order recursive filtering [35] as given by:

$$P_{YiYj}(f,n) = \lambda \cdot P_{YiYj}(f,n-1) + (1-\lambda) \cdot Y_i(f,n)Y_j^*(f,n) \quad i,j = \{1,2\} \tag{4}$$

where * indicates the complex conjugate operator.

λ is a smoothing parameter usually close to 1. This factor should satisfy the two following constraints:

- For low values, the estimation takes into account speech short term stationarity.
- For higher values, this parameter serves to minimize the estimator variance.

Previous study [2] showed that for 16 kHz sampling frequency with 256 samples per frame and a 50% overlap, the upper limit values of λ is around 0.6–0.8. The value $\lambda = 0.7$ is adopted for this study.

In the coherence based methods, the frequency components of the filter vary according to the amount of coherence between channels. It is assumed that the signal source is close to the microphones; and as a result, the received speech signals are correlated. It is also assumed that the distance between adjacent microphones is not too small (in our case, due to the size of the human head, the distance inter-microphones was about 180 mm); consequently, received noise signals can be uncorrelated. Thus, higher values of coherence function correspond to an increased level of desired speech in the signal. Nevertheless, low values of the coherence function correspond to increased level of noise in the signal. These properties led some researchers to use coherence magnitude or a function of it as a spectral modification filter. However, the assumption of uncorrelated received noise is often

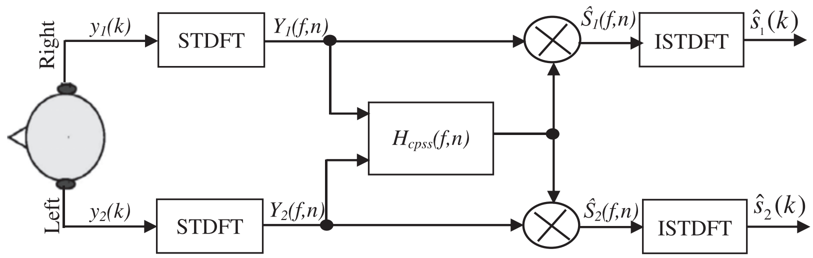

Fig. 1. Block diagram of the considered dual-channel speech enhancement algorithm.

violated in realistic conditions. The CPSS algorithm is considered and the modification filter (H_{cpss}) is computed according to Eq. (5) [34]:

$$H_{cpss}(f,n) = \frac{|P_{Y1Y2}(f,n)| - |P_{D1D2}(f,n)|}{\sqrt{P_{Y1Y1}(f,n)P_{Y2Y2}(f,n)}} \tag{5}$$

where $P_{D1D2}(f,n)$ is the noise CPSD.

A precise estimation of the noise CPSD ($|P_{D1D2}(f,n)|$) is crucial to obtain an accurate estimation of the speech signal. Different techniques are available to compute the noise CPSD like VAD and MT approaches witch will be detailed in the next subsections.

2.2. Noise CPSD estimation using VAD approach

A technique for estimation of the noise CPSD is to employ a voice activity detector (VAD). In a VAD-based technique, the estimation is updated in non-speech (pause) regions and is stopped during speech activity. This is expressed in:

$$\left|\hat{P}_{D1D2}(f,n)\right| = \begin{cases} \lambda'.\left|\hat{P}_{D1D2}(f,n-1)\right| + (1-\lambda').\left|Y_1(f,n).Y_2^*(f,n)\right| & \text{pause} \\ \left|\hat{P}_{D1D2}(f,n-1)\right| & \text{speech frame} \end{cases} \tag{6}$$

where $\hat{P}_{D1D2}(f,n)$ is the estimate of the noise CPSD and λ' is a smoothing parameter and it was set to 0.9 in this work.

Either single or dual microphone VADs can be used to distinguish the speech/pause regions. In this study, we used a coherence-based VAD [1,2] in which the speech/pause regions are determined by taking a threshold on the coherence magnitude.

2.3. Noise CPSD estimation using minimum tracking (MT) approach

Martin [37] proposed a single-microphone noise estimation method based on the observation of the noise PSD estimate witch can be obtained using minimum values of the smoothed power estimate of the noisy speech signal. Thus, the use of minimum statistics eliminates the problem of speech activity detection. Rahmani et al. [24] extended this algorithm to the dual-microphone case for noise CPSD estimation. First, a smoothed noisy CPSD, $R_{Y1Y2}(f,n)$, is computed according to:

$$R_{Y1Y2}(f,n) = \lambda' R_{Y1Y2}(f,n-1) + (1-\lambda')\cdot Y_1(f,n)Y_2^*(f,n) \tag{7}$$

In the case of small signals (such as noise), $R_{Y1Y2}(f,n)$ has low values. The short-time noise CPSD is estimated by finding the local minima ($R_{Y1Y2\min}(f,n)$) of the smoothed noisy CPSD. These values are found on the short intervals of 'V' frames as follows:

$$R_{Y1Y2\min}(f,n) = \min\{|R_{Y1Y2}(f,m)|, n-V+1 < m < n\} \tag{8}$$

where V is the length of the search window and is set to the maximum lifetime of the speech components.

Since the short-term minimum CPSD magnitude is always smaller than the mean CPSD magnitude, the local minima of $|R_{Y1Y2}(f,n)|$ is a biased estimate of the CPSD magnitude that decreases the noise CPSD estimation accuracy. For an accurate noise CPSD estimate, this bias must be compensated. For this reason, it is multiplied by a value named the 'bias compensation factor', B_{min}. The estimated noise CPSD is computed according to:

$$\hat{P}_{D1D2}(f,n) = R_{Y1Y2\min}(f,n)B_{\min} \tag{9}$$

In the case of single-channel noise PSD estimation, Le Bouquin-Jeannès et al. [34] used a constant bias compensation factor to correct the error between the estimated and the true noise PSD. Martin [37] showed that the bias compensation factor is a function of the search window length, V, and the noise PSD estimate variance. Here, B_{min} is set to be a constant value [24] for a fixed search window length ($B_{min} = 2$).

3. Proposed noise CPSD estimation using improved MT (IMT) approach

In MT approach, the noise update was dependent on the length of the minimum-search window. The main drawback of this method is that it takes more than the duration of the minimum-search window to update the noise spectrum when the noise floor increases abruptly. To overcome this problem, Farsi [9] proposed a single-microphone non-linear rule for tracking the minimum of the noisy speech by continuously averaging past spectral values. In this algorithm, the update of local minimum was continuous over time and did not depend on some fixed window length. We proposed then an extension of this algorithm to the dual-microphone case for noise CPSD estimation where the local minima are computed as given by:

$$R_{Y1Y2\min}(f,n) = \begin{cases} \alpha\cdot R_{Y1Y2\min}(f,n-1) + \frac{1-\alpha}{1-\beta}[R_{Y1Y2}(f,n) - \beta\cdot R_{Y1Y2}(f,n-1)] & \text{if } R_{Y1Y2\min}(f,n-1) < R_{Y1Y2}(f,n) \\ \alpha\cdot R_{Y1Y2\min}(f,n-1) + \gamma[\lambda\cdot R_{Y1Y2}(f,n) - R_{Y1Y2\min}(f,n-1)] & \text{otherwise} \end{cases} \tag{10}$$

where α, β, γ and λ are constants experimentally determined by Farsi [9] ($\alpha = 0.997$, $\beta = 0.8$, $\gamma = 0.01$, $\lambda = 0.9$).

The block diagram of the noise reduction algorithm based on IMT approach for noise CPSD estimation is shown in Fig. 2. In this block diagram, the PSD and CPSD values are estimated using Eq. (4), and the smoothed CPSD is then calculated using Eq. (7). In the next step, the noise CPSD is computed using the minima tracking approach using both Eqs. (9) and (10). The estimated noise CPSD is exploited to calculate the noise reduction filter as in Eq. (5). Finally, this filter is applied to the noisy speech signals of both

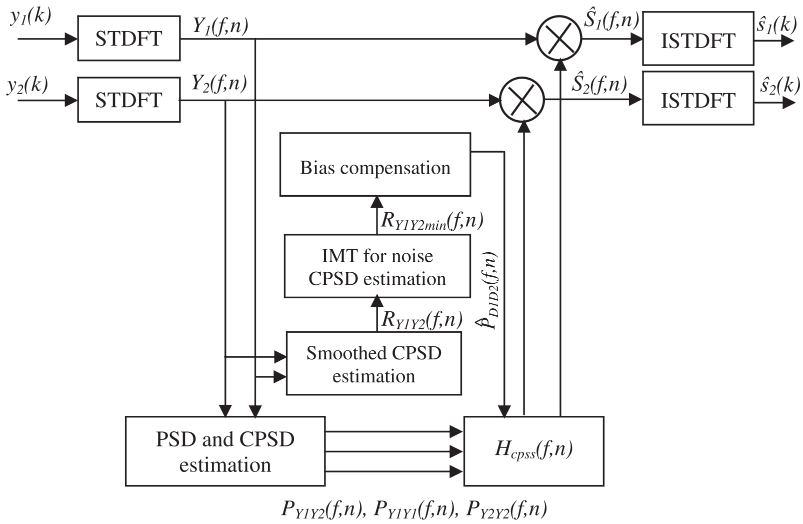

Fig. 2. Block diagram of the considered noise CPSD estimator. $y_1(k)$ and $y_2(k)$ are the input noisy signals. $\hat{s}_1(k)$ and $\hat{s}_2(k)$ are the output enhanced signals.

left and right channels. The modified spectrums ($\hat{S}_1(f, n)$ and $\hat{S}_2(f, n)$) are used to reconstruct the enhanced speech time signals by using the ISTDFT in conjunction with the overlap and add method.

Algorithm 1 indicates the different steps used to compute the enhanced speech signals with the proposed noise CPSD estimation method. As can be seen, the initialization of recursive items is performed in the first frame. Practically, it is assumed that the first frame contains only noise, thus the CPSD of noise is set to the CPSD of the noisy signals, and the modification filter is set to zero (to attenuate all noise).

Algorithm 1. The algorithm of the proposed method for noise CPSD estimator is given below. Recommended values for smoothing parameters are: $\lambda = 0.7$, $\lambda' = 0.9$, $\alpha = 0.997$, $\beta = 0.8$, $\gamma = 0.01$, and $B_{min} = 2$ [9].

First frame (n = 1)
$P_{Y1Y1}(f, 1) = Y_1(f, 1)\cdot Y_1(f, 1)^*$
$P_{Y2Y2}(f, 1) = Y_2(f, 1)\cdot Y_2(f, 1)^*$
$P_{Y1Y2}(f, 1) = Y_1(f, 1)\cdot Y_2(f, 1)^*$
$R_{Y1Y2}(f, 1) = Y_1(f, 1)\cdot Y_2(f, 1)^*$
$R_{Y1Y2min}(f, 1) = R_{Y1Y2}(f, 1)$
$H_{cpss}(f, 1) = 0$

Other frames ($n > 1$)
$P_{Y1Y1}(f, n) = \lambda\cdot P_{Y1Y1}(f, n-1) + (1-\lambda)\cdot Y_1(f, n)Y_1(f, n)^*$
$P_{Y2Y2}(f, n) = \lambda\cdot P_{Y2Y2}(f, n-1) + (1-\lambda)\cdot Y_2(f, n)Y_2(f, n)^*$
$P_{Y1Y2}(f, n) = \lambda\cdot P_{Y1Y2}(f, n-1) + (1-\lambda)\cdot Y_1(f, n)Y_2(f, n)^*$
$R_{Y1Y2}(f, n) = \lambda'\cdot R_{Y1Y2}(f, n-1) + (1-\lambda')\cdot Y_1(f, n)Y_2(f, n)^*$
Non-linear rule for local minima ($R_{Y1Y2min}(f, n)$) computing according to Eq. (10)
$\hat{P}_{D1D2}(f,n) = B_{\min}\cdot R_{Y1Y2\min}(f,n)$
$H_{cpss}(f,n) = \frac{|P_{Y1Y2}(f,n)|-|\hat{P}_{D1D2}(f,n)|}{\sqrt{P_{Y1Y1}(f,n)\cdot P_{Y2Y2}(f,n)}}$
$\hat{S}_i(f,n) = Y_i(f,n)\cdot H_{cpss}(f,n) \quad i = \{1,2\}$

4. Bilateral cochlear implant simulator

In order to investigate the behaviour of the above speech enhancement algorithms, in the case of cochlear implant, speech signals were processed and modified according to a specific cochlear implant simulator leading to vocoded speech signals. Fig. 3 gives the block diagram of the considered CI simulator. The conception of this simulator was based on the French 12-channel Neurelec Digisonic SP cochlear implant. According to several authors [30], these simulations provide results consistent with the outcome of cochlear implants.

The enhanced speech signal, sampled at 16 kHz, was divided into 50% overlapping frames by the application of a Hanning window function and then analysed using the STDFT. The length of the window (frame) is set to $L = 128$ samples. The processed speech signal is passed through a bank of 12 band-pass filters. The bandwidth of the filters matched the frequency range corresponding to the cochlear segment excited by electrical stimulation. It can be noticed that cochlear tonotopy is organized according to psychoacoustic evidence, the lin-log scale (linear in low frequencies and logarithmic in high frequencies) such as the bark scale. In this work, cut-off frequencies between the different filters were chosen in order to respect this scale.

Mapping between the bark scale and the frequency scale is non-linear. An approximate analytical expression for describing the conversion from linear frequency f (in Hz), into the critical band number b (in barks) is given by the following equation [10]:

$$b(f) = ArgHyperbolic\left(\frac{f-20}{600}\right) \tag{11}$$

The number of spectrum lines (N_z), the starting line ($nstart_z$), the center frequency ($fmid$) and low and high cut-off frequencies for each frequency band are summarized in Table 1.

Filter outputs are then processed using the following power estimation equation:

$$E(z) = \frac{\sum_{f=nstart_z}^{f=nstart_z+N_z-1} e(f)}{N_z}, \quad z = 1:12 \tag{12}$$

where $E(z)$ is the power of the considered band 'z' and $e(f)$ is the power of the spectral line 'f'.

The output of this analysis stage is a vector of power values for each data frame. According to the advance combination encoder (ACE) strategy [6] and for each frame, only the first 8 channels presenting the most important power levels were used. The other channels were set to zero.

Each frequency band 'z' was weighted by a Hanning window '$w(l)$' and the energy value '$E(z)$' yielded a modified Hanning window called "the envelope: $Env(z, l)$", according to:

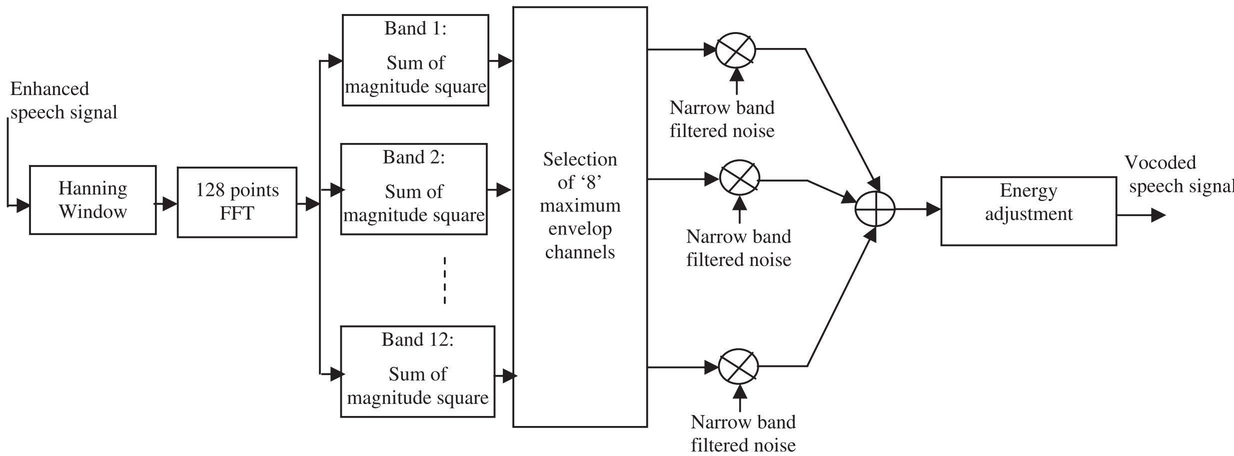

Fig. 3. Block diagram of the adopted CI simulator.

Table 1
Spectral lines attributed to the frequency bands and their associated center frequencies in Hz.

Analysis Frequency band 'z'	1	2	3	4	5	6	7	8	9	10	11	12
Left and right channels												
Number of lines N_z	1	1	2	2	2	3	3	4	5	6	8	9
Starting line $n\text{start}_z$	3	4	5	7	9	11	14	17	21	26	32	40
Center freqs f_{mid} (Hz)	375	500	687	937	1187	1500	1875	2312	2875	3562	4437	5500
Cut-off frequencies (Hz)	300–448	448–616	616–810	810–1041	1041–1318	1318–1654	1654–2065	2065–2568	2568–3187	3187–3950	3950–4892	4892–6055

$$Env(z,l) = E(z) \cdot w(l), z = 1:12, \quad l = 1:L \tag{13}$$

To prevent sharp variations, a further smoothed envelop was obtained after a low pass filtering (cut-off at 150 Hz). A white noise was shaped to each frequency band by a 3rd order Butterworth filter having the same bandwidth than the frequency band, and yielded narrow band noise signals. For each frequency band, a vocoded speech signal was synthesized by modulating the filtered narrow band noise signal with its corresponding smoothed envelop. Finally, all synthesized vocoded signals coming from the different channels were summed up and the output was adjusted to have the same long-term root mean square energy as the input speech signal (70 dB).

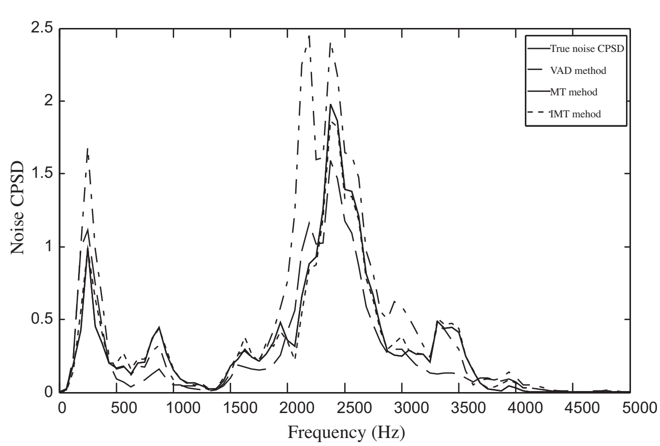

Fig. 4. Performance of the noise CPSD estimators using MT, IMT and VAD methods.

5. Evaluation and results

In this section, we evaluate the performance of previously described approaches (VAD, MT and IMT) for noise CPSD estimation. A preliminary comparative study is proposed to show the usefulness of the proposed approache in terms of better noise CPSD estimation and processing time. Objective measurements and subjective listening tests were then considered. Enhanced speech signals are computed according to the noise reduction filter given by Eq. (5) and using in each case a noise CPSD estimator.

5.1. Preliminary comparative study

Fig. 4 shows the variation of the noise CPSD as a function of the frequency. Noise CPSD was estimated in the presence of speech signal and is obtained by taking the average over all frames. A babble noise with three interfering noise sources was considered. The SNR of the considered noisy speech signal was fixed at 0 dB. True noise CPSD is depicted as a solid line. The estimated noise CPSD is sketched, in dashed-dot line for VAD method, in dashed line for MT method and in dotted line for the proposed IMT method. The length of the search window was fixed to 30 frames (0.48 s) for the MT method.

As it can be noticed, the IMT method performed better than MT and VAD methods and particularly in mid range frequencies (1500 Hz–3500 Hz). In fact, the noise CPSD estimated with the IMT method fit well the true noise CPSD.

These methods were also compared in term of processing time using AMD 2.4 GHz clock processor machine and 1Go of RAM under windows operating system. Different search windows length (V = 10, 20, 30, 40, 50 and 60 frames) for the MT algorithm were considered. Results are reported in Table 2.

Table 2
Processing time for MT, IMT and VAD methods.

Method	Length of search window (frames)	Processing time (second)
MT	10	1.04
	20	1.35
	30	1.55
	40	1.82
	50	2.06
	60	2.32
VAD	–	1.22
IMT	–	1.13

Table 2 shows that the processing time for noise CPSD estimation obtained by the MT approach was increased with the length of the search window. The IMT approach gave a small processing time, compared to MT and VAD methods.

5.2. Intelligibility evaluation

5.2.1. Phonetic material

Phonetic material was the French Lafon set [12] which contains twenty lists, each one composed of seventeen 3-phoneme words pronounced by a single male talker. This set is commonly used for intelligibility assessment in French. As indicated above, sound level was calibrated to 70 dB SPL which is a comfortable hearing level. All theses lists were processed in the anechoic room of the ORL department of the Edouard-Herriot Hospital of Lyon-France with an additional babble noise. This room is fitted with six loudspeakers (LS) and a Kemar (Knowles Electronic Manikin for Electronic Research). SNR was varied from −3 dB to 6 dB with 3 dB steps.

The experimental setup is presented in Fig. 5. A CD player (PHILIPS-CD723) was connected to an audiometer (MADSEN-Orbiter 922) for calibration and intensity level adjustment. The target speech signal was always placed directly in front of the Kemar head at 0° azimuth (LS_3). Subjects were tested with either one or three interferers. In the single interferer condition, a single speech babble noise source was presented from the right side of the listener (LS_5). In the multiple interferers condition, three interfering noise sources were placed asymmetrically either across both hemifields (−60°, 60°, and 90° corresponding respectively to LS_2, LS_4 and LS_5). Signal was recorded, processed and then stored before its delivery to the subjects.

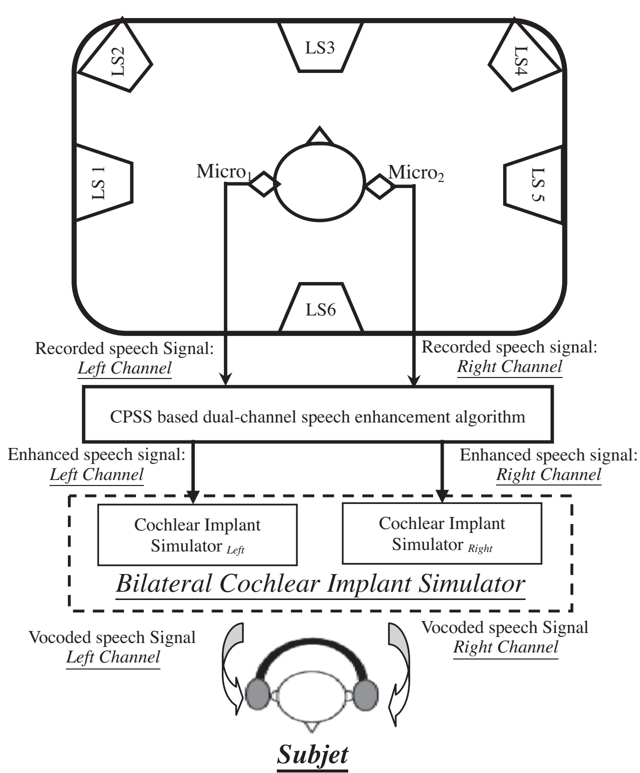

Fig. 5. Experimental setup (Anechoic room). LS_3 = speech signal, LS_2, LS_4 and LS_5 = noise sources, LS_1 and LS_6 were not used.

5.2.2. Objective evaluation

The performance of the different described noise CPSD estimators was evaluated using the PESQ (perceptual evaluation of speech quality) score which range from 0.5 (for the worst case) to 4.5 (for the best case) according to the ITU-T Recommendation P. 862 standard [11]. The results were reported for four SNRs and two interfering noise source configurations. From Table 3, it is clear that in most cases, PESQ scores obtained with minimum tracking approaches were higher than those of VAD based method. This can be explained by the fact that VAD did not take into account the behaviour of the noise during the speech frames, while the minimum tracking approaches (MT and IMT) estimated the noise CPSD in sub-bands during speech and non-speech frames.

It can be also noted that PESQ values obtained with IMT approach were higher than those for the MT. Furthermore, the worse PESQ scores were obtained at lower SNR levels and with multiple interfering noise sources (second configuration).

5.2.3. Subjective evaluations

In this section, we investigate the potential benefits of processing the noisy speech signal with the CPSS algorithm using three different noise CPSD estimators. Therefore, comprehensive phoneme recognition tests were conducted in BCI simulation.

5.2.3.1. Listeners. Performance evaluation was carried out with a population of fifty normal hearing subjects. Their audiograms were tested in the ORL department prior to the experiment. Normal hearing subjects age ranged from 20 to 35 years. All participants were native French speakers. Tests were performed in the anechoic room of the Edouard-Herriot hospital of Lyon.

5.2.3.2. Listening session. After listening to each word, subjects were instructed to repeat what they heard. Before each condition, subjects were given a practice session containing a set of ten random words processed according to that condition. None of the words used during the test was used in the practice session. No score was calculated for these practice sets. To minimize any order effect in the experience, such as learning or fatigue, all conditions were randomized among subjects. Different sets of lists were used in each condition. At the end of each listening session, the responses of each individual were collected, stored and scored off-line with the number of correctly identified phonemes. All phonemes were scored. The percentage of correctly repeated phonemes was then calculated (for two Lafon's lists, corresponding to 102 phonemes). Experiments were performed using a PC equipped with a conexant AC-link audio soundcard.

Subjective evaluation tests were elaborated considering noisy speech signal (non-processed signal) taken as a reference condition and three SEAs based on CPSS using VAD, MT and IMT approaches for noise CPSD estimation. Speech stimuli were processed offline with MATLAB software. Each subject listened to a total of 64 lists (8 lists/condition × 4 SNRs × 2 noise interfering configurations). Lists were played on the CD player in a random order. For each condition, the subjects listened bilaterally to the words using a closed professional 'Sennheiser' HD250 linear headphones. Acoustic level was calibrated at 70 dB SPL.

Table 3
PESQ scores of noisy and enhanced speech signals.

Method	*PESQ*							
	Single interfering noise source				Three interfering noise sources			
	SNR = −3 dB	SNR = 0 dB	SNR = 3 dB	SNR = 6 dB	SNR = −3 dB	SNR = 0 dB	SNR = 3 dB	SNR = 6 dB
Noisy signal	1.23	1.62	1.98	2.08	1.13	1.36	1.52	1.94
VAD	1.32	1.84	2.26	2.71	1.06	1.70	2.09	2.31
MT	1.86	2.13	2.44	2.62	1.52	1.96	2.28	2.48
IMT	2.04	2.31	2.58	2.79	1.77	2.04	2.40	2.62

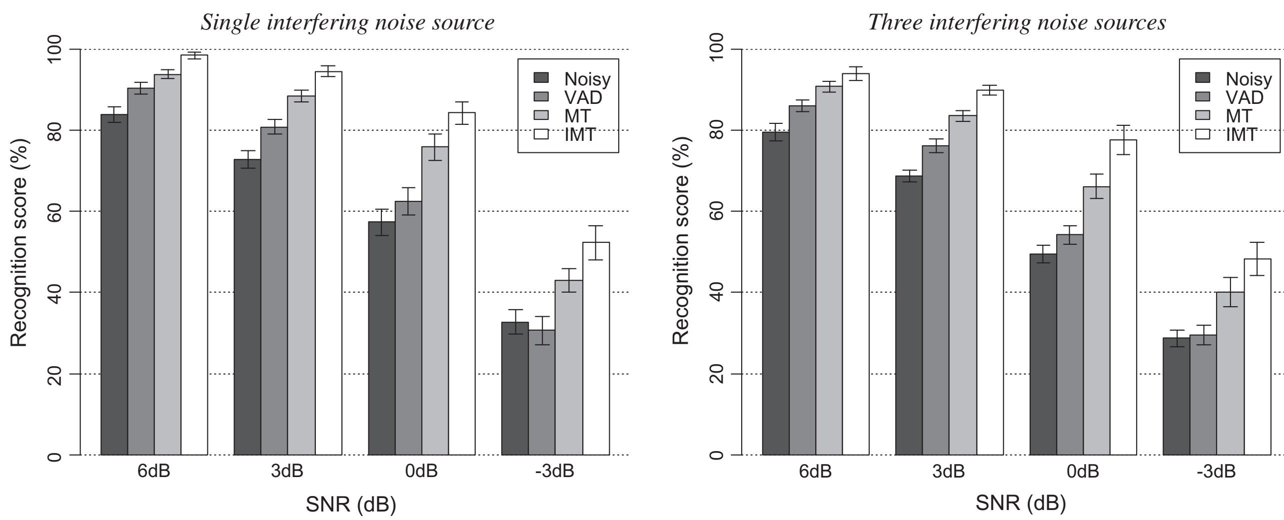

Fig. 6. Recognition score (mean ± standard error) in% with noisy speech, CPSS based SEA using VAD, MT and IMT noise CPSD estimator for two Babble noise interfering source configurations.

5.2.3.3. Results. Performance of CPSS speech enhancement algorithm with VAD, MT and IMT noise CPSD estimation methods were carried out through listening tests. Intelligibility scores were derived from the percentage of correctly repeated phonemes per condition. Effects were studied through Chi2 tests with the following parameters:

- Repeated measures (the same subjects underwent all the situations)
- Dependant variable: the recognition score in percent
- Three effects:
  - SEA (Noisy, VAD, MT and IMT)
  - Interfering noise configuration ('90°' and '−60° 60° 90°')
  - SNR (−3 to 6 dB with 3 dB step).

This last effect was taken as random; the first two effects were fixed. This model, with both fixed and random effects, is known as mixed-effect model, and is a standard for data analysis [41]. Analysis was carried out using the program lmer of the lme4 package [7] of the statistical programming language and software environment R [36]. Results revealed main effects of SEA (Chi2[3] = 1244, $p < 0.0001$), of SNR (Chi2[3] = 6461, $p < 0.0001$) and of interfering noise configuration (Chi2[1] = 121, $p < 0.0001$). In addition, there was a significant interaction between the SEA and the SNR (Chi2[9] = 77, $p < 0.0001$). No significant interaction was seen between interfering noise configuration and SEA (Chi2[3] = 8, $p = 0.03$) and between interfering noise configuration and SNR (Chi2[3] = 8, $p = 0.2$).

Fig. 6 shows mean recognition scores (across all subjects) in the presence of speech babble noise, as a function of the SNR for both considered noise interfering configurations. It is clear that performance of CPSS based SEA is variable with the considered approach for noise CPSD estimation. An overall superiority of the proposed technique compared to the others is observed.

Post-hoc comparisons were run to assess significant differences. We used the Tukey HSD test via general linear hypothesis test '*glht*' function from multiple comparison '*multcomp*' package of R. Results revealed that phoneme recognition scores with VAD approach (particularly at SNRs 6, 3 and 0 dB) were significantly better than the unprocessed noisy speech signal ($p < 0.001$), but no significant improvement was seen at −3 dB ($p = 0.08$). However, best recognition scores were obtained with MT and IMT at all SNRs ($p < 0.001$). Results indicated also that phoneme recognition scores with MT and IMT were significantly better than those obtained with VAD at all SNRs ($p < 0.001$). Furthermore, it can be noted that IMT performed significantly better than MT at 3, 0 and −3 dB SNRs ($p < 0.001$). No significant improvement was seen at 6 dB ($p = 0.142$).

5.3. Discussion

In the present study, benefits of CPSS SEA using three noise CPSD estimators were assessed. The performance of VAD and MT approaches were compared to the IMT approach. Simulation results showed lower computing time and a better quality in terms of PESQ scores for the enhanced speech signal when the proposed noise CPSD estimator was considered. Moreover, subjective comparative study based on the phoneme recognition scores was also performed with fifty normal-hearing subjects. Globally, experimental results indicated an overall superiority of the IMT method.

5.3.1. Speech enhancement algorithms vs SNR levels

Results showed a variable effect of SEAs when SNRs were decreased from 6 dB to −3 dB. In fact, the average improvements in phoneme's recognition scores were 5.2%, 13.6% and 20.1% obtained respectively with VAD, MT and IMT approaches. Results also demonstrated that the improvement was fairly high at 6, 3 and 0 dB, but was less at −3 dB SNR. These results indicate the superiority of minimum tracking approaches for noise CPSD estimation, and IMT was better than MT and VAD.

Compared to the mean recognition scores obtained with multiband spectral subtraction algorithm proposed by Kallel et al. [8], an average improvements of 3%, 4.5%, 8.5% and 6.5% were respectively observed at 6 dB, 3 dB, 0 dB and −3 dB when the CPSS SEA were considered using the IMT approach for noise CPSD estimation.

5.3.2. Speech enhancement algorithms vs interfering noise source configurations

As shown in Fig. 6, SEAs performances were significantly degraded when the number of interfering noise sources was increased from one to three. In the presence of only one interfering noise source, the average improvement was 6% with VAD, 13.5% with MT and 24.6% with IMT. In the presence of three interfering noise sources, results indicated an average improvement of only 4.5% with VAD, 10.7% with MT and 21.2% with IMT.

This observation may be due to the fact that, in the binaural hearing case, when both CI devices are available, BCI listeners could benefit from the head-shadow effect which occurs mainly when speech and noise are spatially separated [27,40]. When only one interfering noise source was placed on the right of the listener (+90°), a better SNR was obtained at the left ear compared to the right one. Therefore, the individual listener will be able to selectively focus on the left ear placed contralateral to the competing noise source to improve phoneme recognition scores. Whereas for multiple interfering noise sources, the speech signals coming to the right and left sides of the listener will be disproportionally attenuated depending on the noise source configuration.

The implementation of the proposed dual-channel speech enhancement strategy requires access to a single digital signal processor driving both cochlear implants. In this case, signals coming from the left and right sides need to be captured synchronously and be further processed together. As wireless communication of speech signals is quickly becoming possible between the two sides, it may be advantageous to have the two devices in a bilateral fitting communicate with each other. An intelligent wireless signal transmission scheme could also be designed that would allow a signal exchange between the two implant processors.

6. Conclusion

This study investigated the efficiency of CPSS based two-microphone SEA applied to BCI case in simulation. The noise CPSD estimation was an important component. An advanced approach for noise CPSD estimation, based on improved minimum tracking technique, was presented. Tests were performed using the French Lafon set based on phoneme recognition. Speech babble was the interfering noise signal. Fifty normal hearing subjects participated in the experiment. Performances of noise CPSD estimators were evaluated to compare VAD and MT approaches in single and three interfering noise sources at different SNRs. Objective measures based on PESQ scores indicated a significant improvement in speech quality when the IMT approach was considered. A minimum computing time was also observed. Subjective recognition tests revealed that CPSS algorithm can yield substantial benefits in speech intelligibility for BCI in simulation when the IMT approach for noise CPSD estimation is considered, especially in the presence of only one interfering noise source.

Acknowledgments

This research work was supported by the French–Tunisian cooperation project 'PHC-Utique' managed by CMCU (2008–2011). We appreciate the efforts of all the subjects who participated in the listening sessions. We also would like to thank the cochlear implantation center in the Edouard Herriot hospital in Lyon-France for the technical support and especially professors Lionel Collet, Hung Tai-Van and Eric Truy for their encouragements.

References

[1] A. Guérin, Rehaussement de la parole pour les communications mains-libres: Réduction de bruit et annulation d'echo non linéaire, PhD thesis, Université de Rennes 1, France; 2002.
[2] Guérin A, Le Bouquin-Jeannès R, Faucon G. A two-sensor noise reduction system: applications for hands-free car kit. EURASIP J Appl Signal Process 2003:1125–34.
[3] Spahr A, Dorman M. Performance of patients fit with advanced bionics CII and nucleus 3G cochlear implant devices. Arch Otolaryngol Head Neck Surg 2004;130:624–8.
[5] Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble1 W. Bilateral and unilateral cochlear implant users compared on speech perception in noise. Ear Hear 2010;31:296–8.
[6] Cochlear Corporation. ACE and CIS DSP strategies. Software requirements specification. Lane Cove, New South Wales, Australia; 2002.
[7] Bates D. lme4: linear mixed-effects models using S4 classes. R package version 0.99875-8; 2007.
[8] Kallel F, Frikha M, Ghorbel M, Ben Hamida A, Berger-Vachon C. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. Appl Acoust 2012;73:12–20.
[9] Farsi H. Improvement of minimum tracking in minimum statistics noise estimation method. Signal Process Int J 2010;4:17–22.
[10] Tranmüller H. Analytical expression for the tonotopic sensory scale. J Acoust Soc Am 1990;85:97–100.
[11] ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs; 2001.
[12] Lafon JC. Le test phonétique et la mesure de l'audition. In: Eindhoven, editors. Centrex; 1964.
[13] Deller JR, Hansen JHL, Proakis JG. Discrete-time processing of speech signals. 2nd ed. New York: IEEE Press; 2000.
[14] Wouters J, Vanden Berghe J. Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system. Ear Hear 2001;22:420–30.
[15] Chung K, Zeng FG, Acker KN. Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. J Acoust Soc Am 2006;120:2216–27.
[16] Kasturi K, Loizou PC. Use of S-shaped input-output functions for noise suppression in cochlear implants. Ear Hear 2007;28:402–11.
[17] Reindl K, Zheng Y, Kellermann W. Analysis of two generic wiener filtering concepts for binaural speech enhancement in hearing aids. In: 18th European signal processing conference; 2010. p. 989–93.
[18] Kokkinakis K, Loizou PC. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. J Acoust Soc Am 2010;127:3136–44.
[19] Yang LP, Fu QJ. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. J Acoust Soc Am 2005;117:1001–4.
[20] Monica LH, Ruth YL. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J Acoust Soc Am 2004;115:833–43.
[21] Kompis M, Dillier N. Noise reduction for hearing aids: combining directional microphones with an adaptive beamformer. J Acoust Soc Am 1994;96:1910–3.
[22] Rahmani M, Akbari A, Beghdad A. An iterative noise cross-PSD estimation for two-microphone speech enhancement. J Appl Acoust 2009;70:514–21.
[23] Rahmani M, Akbari A, Ayad B, Lithgow B. Noise cross PSD estimation using phase information in diffuse noise field. J Signal Process 2009;89:703–9.
[24] Rahmani M, Akbari A, Ayad B, Derakhshan N. A noise cross PSD estimator for dual-microphone speech enhancement based on minimum statistics. J Zhejiang Univ Sci A 2009;10:805–9.
[25] Qin MK, Oxenham AJ. Effects of simulated cochlear implant processing on speech reception in fluctuating maskers. J Acoust Soc Am 2003;114:446–54.
[26] Kompis M, Bettler M, Vischer M, Senn P, Häusler R. Bilateral cochlear implantation and directional multi-microphone systems. In: International cochlear implant conference 1273; 2004. p. 447–50.
[27] Hawley ML, Litovsky RY. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J Acoust Soc Am 2004;115:833–43.
[28] Rahmani M, Yousefian N, Akbari A. An energy-based speech enhancement technique for hands-free communication. Electron Lett 2009;45:85–6. IEE Publishing.
[29] Yousefian N, Kokkinakis K, Loizou PC. A coherence-based algorithm for noise reduction in dual-microphone applications. In: European signal processing conference 1904–1908; 2010.

[30] Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. J Acoust Soc Am 2007;122:2376–88.
[31] Loizou PC, Hu Y, Litovsky P, Yu G, Peters R, Lake J, et al. Speech recognition by bilateral cochlear implant users in a cocktail party setting. J Acoust Soc Am 2009;125:372–83.
[32] Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. J Acoust Soc Am 2005;118:2791–3.
[33] Sovka P, Pollak P, Kybic J. Extended spectral subtraction. In: Proceeding of European signal processing conference, EUSIPCO-96. Trieste, Italia; 1996.
[34] Le Bouquin-Jeannès R, Azirani AA, Faucon G. Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator. IEEE Trans Speech Audio Process 1997;5:484–7.
[35] Le Bouquin-Jeannès R, Faucon G. Using the coherence function for noise reduction. IEE Proc 1992;139:276–80.
[36] R Development Core Team, R: A language and environment for statistical computing, R foundation for statistical computing. Vienna, Austria; 2007. ISBN 3-900051-07-0. <http://www.R-project.org>.
[37] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans Speech Audio Process 2001;9:504–12.
[38] Martin R. Spectral subtraction based on minimum statistics. In: Signal processing conf. EUSIPCO 1182–1185; 1994.
[39] Van Hoesel RJM, Clark G. Evaluation of a portable two microphone adaptive beamforming speech processor with cochlear implant patients. J Acoust Soc Am 1995;97:2498–503.
[40] Litovsky RY, Parkinson A, Arcaroli J. Spatial hearing and speech intelligibility in bilateral cochlear implant users. Ear Hear 2009;30:419–31.
[41] Baayen RH. Analyzing linguistic data. A practical introduction to statistics. Cambridge, UK: Cambridge University Press; 2008.
[42] Hamacher V, Doering W, Mauer G, Fleischmann H, Hennecke J. Evaluation of noise reduction systems for cochlear implant users in different acoustic environments. Am J Otol 1997;18:S46–9.
[43] Noble W, Tyler R, Dunn C, Bhullar N. Hearing handicap ratings among different profiles of adult CI users. Ear Hear 2008;29:112–20.
[44] Zhang X, Jia Y. A soft decision based noise cross power spectral density estimation for two-microphone speech enhancement systems. In: IEEE int conf on acoustics, speech, and signal processing; 2005. p. 813–816.
[45] Hu Y, Loizou P, Li N, Kasturi K. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. J Acoust Soc Am 2007;122:128–34.

Applied Acoustics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Influence of a shift in frequency distribution and analysis rate on phoneme intelligibility in noisy environments for simulated bilateral cochlear implants

Fathi Kallel [a,b,*], Rafael Laboissiere [b], Ahmed Ben Hamida [a], Christian Berger-Vachon [b]

[a] Research Unit in Advanced Technologies for Medicine and Signals 'ATMS', National Engineering School of Sfax, University of Sfax, Route Soukra, km 3, Sfax, BPW 3038, Tunisia
[b] PACS Team, INSERM Unit 1028: "Cognition and Brain Dynamics", Lyon Neurosciences Research Centre, EPU-ISTIL, Claude Bernard University, Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France

A R T I C L E I N F O

Article history:
Received 11 December 2010
Received in revised form 12 May 2012
Accepted 28 May 2012
Available online xxxx

Keywords:
Bilateral CI stimulation
Analysis rate
Vocoder
Hearing in noise
Asymmetrical stimulation

A B S T R A C T

A model was developed to simulate acoustically sound perception through a cochlear implant (CI), in order to evaluate the effects of a spectral shift and analysis rate on speech recognition in quiet and noise. In the current study, we considered two analysis rates, 250 Hz and 500 Hz, and two CI modes: Symmetric Bilateral Cochlear Implant (SBCI) and Shifted Bilateral Cochlear Implant (ShBCI). Processing and coding strategies used in this model were adapted from the Digisonic SP CI, manufactured by Neurelec.

Intelligibility of speech signals processed to simulate different analysis rates and CI modes were assessed by a group of fifty normal-hearing subjects. The analysis rate was simulated by varying the overlap between successive analysis frames using a narrow band vocoder. With the SBCI mode, both ears were stimulated with the same signal (the same frequency filters were used). With the ShBCI mode, the filters were shifted in frequency (between the two ears). All the conditions were tested in quiet and in noisy environment with three different Signals to Noise Ratios (SNR). The database testing procedure used in this experimentation involved 3-phoneme words selected from the French Lafon's lists. Speech signals were corrupted by addition of speech multi-talker babble noise.

Results showed a significant effect of CI mode, of analysis rate and of SNR. The analysis rate effect was small in quiet and significant in noisy environments. The 500 Hz analysis rate led to better performances than the 250 Hz. Higher performances were also observed with the ShBCI mode in noisy environments. Results were mainly consistent with findings obtained from previous cochlear implant studies which suggest that CI users may perform better with a shifted bilateral stimulation in noisy environments, and with a higher analysis rate.

1. Introduction

Cochlear implants (CI) are electronic devices introduced surgically into the inner ear that directly stimulate the auditory nerve in response to sounds. They are designed for severely, profoundly, or totally hearing-impaired patients who get little or no benefit from classical hearing aids [5,20]. Three main parts can be distinguished in such auditory prosthesis: an external part containing the speech processing analyser which extracts most essential parameters, a transmission module and an internal implanted stimulator [20].

Several speech processing algorithms were elaborated by research groups to extract essential parameters for the cochlea's electric stimulation [20]. This electric stimulation of nerve cells' leads induces a nervous message, transmitted to hearing areas in the brain and lead to subjective interpretation.

The external sound processor decomposes the input audio signal into different frequency bands and delivers energy, in each frequency band, to the appropriate electrode in the cochlea [18]. The influence of the filters has been investigated in several studies such as Kasturi and Loizou [16], Fourakis et al. [12]; they advocated the placement of more filters in the F1/F2 region in order to achieve a better representation of the first two-formant. Filter-spacing, which included log, mel (Fant, 1973) and critical-band [42] spacing, has been widely investigated for the recognition of vowels. Results indicated that many subjects performed better using the critical-band spacing compared to the log spacing. Performance using the mel frequency spacing was lower compared to the two other frequency spacings. The authors attributed this result to the number of frequency bands situated in the F1 and F2 range. When cochlear tonotopy is organized according to a classical lin-log scale (linear in the low frequencies and log in the high

* Corresponding author at: Research Unit in Advanced Technologies for Medicine and Signals 'ATMS', National Engineering School of Sfax, University of Sfax, Route Soukra, km 3, Sfax, BPW 3038, Tunisia. Tel.: +216 74 27 40 88; fax: +216 74 27 55 95.

E-mail address: fathikallel@yahoo.fr (F. Kallel).

0003-682X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2012.05.014

frequencies), named also bark scale, filter cut-off frequencies were chosen in order to respect this scale. Tranmüller [32] made a review of the existing functions describing the tonotopic sensory scale.

The envelope information in each channel is transmitted to the individual electrodes. The acoustic amplitude envelope is used to modulate the electric current delivered by the implanted electrodes. The transmission of acoustic envelope cues is linked to the spectral analysis and to the stimulation rate of the CI device. Better recognition performance could be obtained with high pulse-rates which better code the speech temporal modulation. Also, higher stimulation rates may increase the stochastic response properties of the activated neurons [15]. Despite the theoretical advantages of higher stimulation rates, the outcomes of several studies did not always back this idea [34].

Presently, many researchers are working towards improving perceived speech quality for hearing impaired people. Technology authorises many signal processing strategies in multichannel CIs, and can offer efficient speech understanding [3,8]. Consequently, unilateral cochlear implantation is widely accepted and offers an effective way to restore speech understanding mostly in quiet conditions [13,41]. However, patients encounter difficulties in noisy environments [37,27,25,11].

Bilateral cochlear implantation has been shown to support improved sound localization and lead to better speech perception, particularly in noisy listening conditions. In fact, bilateral hearing offers the opportunity to listen with the ear that has the most favorable signal-to-noise ratio (SNR). Consequently, this is important when speech and noise come from different directions; it uses the head shadow effect and/or the phase difference. Binaural advantage is the ability to combine sounds coming from the two ears which is better than one ear alone [17,26]. For totally deaf patients, it is possible to achieve a binaural stimulation through bilateral cochlear implantation [9]. But, [41] noted that some bilateral listeners "may not have sufficient residual auditory capacity in the central nervous system to make use of binaural cues." Some investigations showed that bilateral CI was beneficial to some individuals in some conditions [7,35]. Several practical studies have shown an advantage for bilateral cochlear implantation in speech-recognition tasks conducted in controlled laboratory settings [33,30]. A comparison of bilateral CI scores and unilateral CI scores in quiet indicated a significant difference between the two situations, the bilateral CI group scoring 19% higher for sentences and 24% higher for words. Laszig et al. [22] examined the benefits of bilateral cochlear implantation in terms of speech recognition in quiet and in noise for hearing-impaired adults using the Nucleus 24 cochlear implant. In noise, better results were observed for the ear closest to the speech source, compared to the other ear near the noise source. Ricketts et al. [27] compared speech recognition, in quiet and at +5, +10, +15, and +20 dB SNRs using five uncorrelated noise sources for bilateral and unilateral modes, in postlingually-deafened adults bilaterally implanted. Results revealed a good bilateral advantage in low SNR conditions. Dunn et al. [6] also compared speech performance in noise with patients wearing bilateral and unilateral CIs. The bilateral CI group showed significantly improved speech perception in noise, compared with unilateral CI subjects.

Few studies have examined the effects on speech perception of spectral asymmetry in relation to the background conditions. The investigation of this situation is important for understanding the performance of CI users in complex environments, and also for exploring the mechanisms underlying speech perception in noise. In the present study, a 12-channel vocoder was taken as a CI simulator. According to several authors [40], this simulation provides results consistent with the outcome of cochlear implants. The "vocoded" speech was presented to normal-hearing listeners in experimental situations. The goal of this study is to compare phoneme recognition scores using two BCI modes (in simulation), in quiet and in the presence of additional speech babble noise at different SNR levels; 6 dB, 0 dB and −6 dB. Bilateral (SBCI and ShBCI) CI modes were also considered: performance of both bilateral CI modes were measured at a low (250 updating per second (ups)) rate and at a moderate (500 ups) rate.

2. Cochlear implant simulator

2.1. Frequency bands (channels) constitution

A series of speech signals were presented to normal hearing listeners. The signals were pre-processed to be representative of the signal perceived by cochlear implantees. This principle was first described by Dudley [10] and then reproduced by several authors (e.g. [31,11]). The speech sounds were processed according to the CI modes and the analysis rates. Tests were realized in quiet and in noise. Different types of stimulation were generated following the literature [21] based on the vocoder principle (Fig. 1).

The cochlear implant simulator was developed using speech processing parameters of the 12-channel Neurelec Digisonic SP CI. Twenty-four channels were used to simulate SBCI and ShBCI CI modes (12 channels × 2 ears). Acoustic signal bandwidth ranged from 300 to 6055 Hz. Several ways of allocating the filters in the frequency domain were considered. Tranmüller's equation [32] was used in the current study:

$$b(f) \approx 6.7 * ArcHyperbolic\sin\left(\frac{f-20}{600}\right) \quad (1)$$

where b is in barks and f is in Hz.

Many CI coding strategies exist, but with few exceptions, they are all variants of the Continuous Interleaving Sampling (CIS) or Advanced Combination Encoder (ACE) methods [4]. They split the input speech signal into short time-segments (frames) and they used a filter bank to yield a M-band spectral representation. 'N' bands ($N < M$ for ACE, $N = M$ for CIS) with the largest amplitudes are selected and compressed in order to match the narrow dynamic range of electrical hearing stimulation. In the current study, the ACE strategy was adopted to simulate BCI, with $M = 12$ and $N = 8$. Practically, the considered frequency range (300–6055 Hz) was divided into 'M' bands. This frequency range spanned from 3 to 20 barks; the spacing step was 1.42 barks for both SBCI and ShBCI modes (Table 1). Sampling rate was 16 kHz. The speech signal, $y(l)$ (l is the discrete-time index), was divided into overlapping frames by the application of a window function. Then a Short-Time Fast Fourier Transform (STFFT) was applied:

$$Y(k,\lambda) = \sum_{n=0}^{N-1} y(l+\lambda\cdot U)\cdot w(l)\cdot e^{-j\frac{2\pi}{L}l\cdot k} \quad (2)$$

where k is the frequency bin index, λ is the time frame index, U is the frame length in time and w is the analysis window (Hanning window) of size L given by the following equation:

$$w(l) = 0.5\left(1-\cos\left(2\cdot\pi\frac{l}{L}\right)\right) \quad (3)$$

The window length (frame) was set to $L = 128$ samples, leading to 64 spectrum bins. FFT bins were then combined to build up the 12 channels.

In the SBCI mode, both right and left sides had the same filters; the number of FFT bands assigned to the left and the right analysis channels, and the center frequencies are given by the upper part of Table 1. With the ShBCI mode, frequency bands were shifted ½ bark up for the right ear (bottom part of Table 1).

The energy of the kth frequency bin is given by the following equation:

Please cite this article in press as: Kallel F et al. Influence of a shift in frequency distribution and analysis rate on phoneme intelligibility in noisy environments for simulated bilateral cochlear implants. Appl Acoust (2012), http://dx.doi.org/10.1016/j.apacoust.2012.05.014

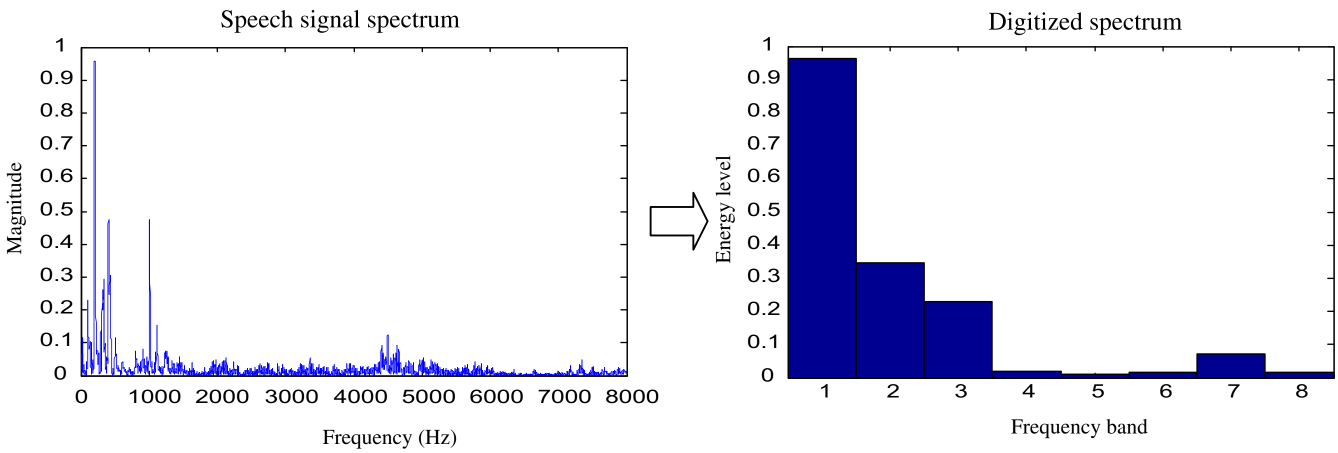

Fig. 1. Spectrum digitization according to CI coding (channel representation).

$$e(k) = Y_r^2(k) + Y_I^2(k) \quad (4)$$

where k is the frequency bin index and $Y_r^2(k)$ and $Y_i^2(k)$ are the real and imaginary parts of the bin.

Then another stage was added to get the signal envelope. The power $E(m)$ of each band was calculated using Eq. (5), based on the Parseval relation.

$$E(m) = \frac{\sum_{k=nstart_m}^{k=nstart_m+N_m-1} e(k)}{N_m}, \; m = 1 : M \quad (5)$$

where m is the frequency band (channel) index. N_m and $nstart_m$ are indicated in Table 1.

Fig. 2 represents the different bands for left and right sides, when the ShBCI mode was considered.

In the ACE strategy, only the first 'N' channels presenting the most important power levels are used for each frame. The remaining channels are set to zero. In the current work, $N = 8$ was taken. To reconstruct the acoustic signal, for each frequency band 'm', a Hanning window '$w(l)$' was weighted by its related power value '$E(m)$' to get the envelope signal, $Env(m, l)$, according to the following equation:

$$Env(m,l) = E(m) \cdot w(l), \quad m = 1 : M, \quad l = 1 : L \quad (6)$$

To prevent sharp variations, a further low pass filtering (cut-off at 150 Hz) was applied to smooth the envelope. Then, a white noise was shaped to fill each frequency band (3rd order Butterworth filter). In each frequency band, the speech signal was synthesized by the multiplication of the shaped narrow band noise signal by the corresponding smoothing envelope. The final test signal was produced by added all the channels together. The level of the final processed speech signal was normalized so that the speech was reproduced with the same Intensity as measured than the original speech signal (70 dB). Fig. 3 sums up the different steps indicated above.

Table 2
Mean recognition percentages according to the different situations.

	Mode	SBCI	ShBCI
Rate	SNR	Mean recognition score %	
250 Hz	NAN	93.3	95.9
	6 dB	81.3	85.4
	0 dB	54.6	58.2
	−6 dB	17	20.7
500 Hz	NAN	95	97.3
	6 dB	84	88
	0 dB	58	65
	−6 dB	20.4	24.9

2.2. Analysis rate

The stimulation rate is linked to the number of pulses per second generated at each electrode of the cochlear implant. This parameter represents the temporal resolution of the implant, which is related to the perception ability of fast changes in the speech signal at high stimulation rates [38,29,19].

Table 1
FFT bins attributed to the frequency channels and their corresponding frequencies in Hz for SBCI and ShBCI.

Frequency Channels 'm'	1	2	3	4	5	6	7	8	9	10	11	12
Left ear												
Number of bins N_m	1	1	2	2	2	3	3	4	5	6	8	9
Starting bin $nstart_m$	3	4	5	7	9	11	14	17	21	26	32	40
Center frequencies f_{center} (Hz)	375	500	687	937	1187	1500	1875	2312	2875	3562	4437	5500
Cut-off frequencies (Hz)	300–448	448–616	616–810	810–1041	1041–1318	1318–1654	1654–2065	2065–2568	2568–3187	3187–3950	3950–4892	4892–6055
Right ear												
Number of bins N_m	2	1	2	2	2	3	4	4	6	7	8	10
Starting bin $nstart_m$	3	5	6	8	10	12	15	19	23	29	36	44
Center frequencies f_{center} (Hz)	437	625	812	1062	1312	1625	2062	2562	3187	4000	4937	6062
Cut-off frequencies (Hz)	365–523	523–704	704–916	916–1168	1168–1473	1473–1843	1843–2297	2297–2854	2854–3540	3540–4387	4387–5432	5432–6735

Please cite this article in press as: Kallel F et al. Influence of a shift in frequency distribution and analysis rate on phoneme intelligibility in noisy environments for simulated bilateral cochlear implants. Appl Acoust (2012), http://dx.doi.org/10.1016/j.apacoust.2012.05.014

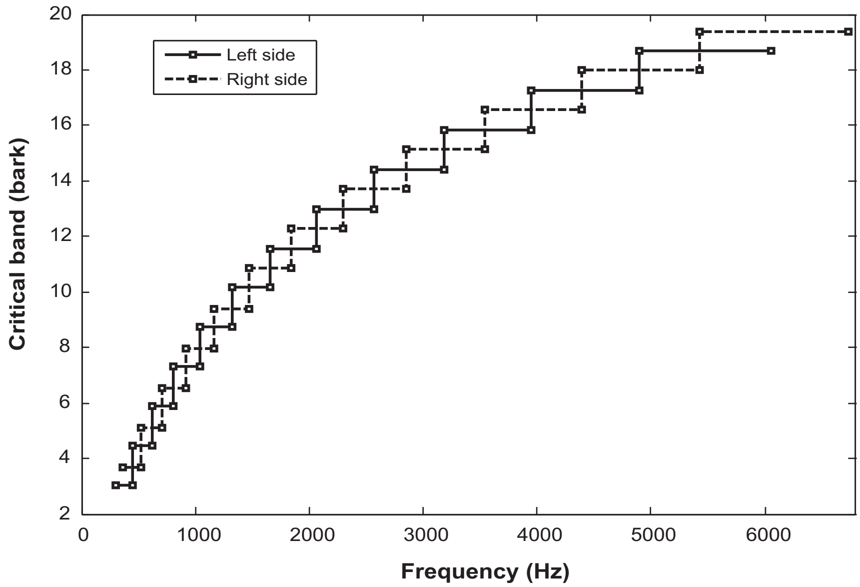

Fig. 2. Low and upper limits of the frequency bands, in Hz and bark, with the ShBCI mode.

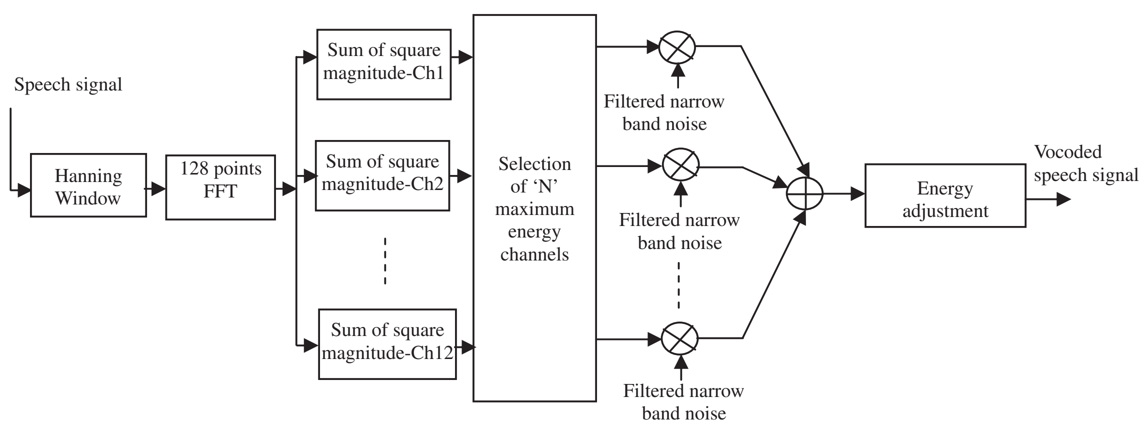

Fig. 3. Block diagram of the CI simulator.

In this study, the stimulation rate has been represented by the analysis rate. Therefore, the value assigned to the parameter "rate" was the updating rate. The analysis rate was controlled by adjusting of the overlap between the successive analysis windows. It will be indicated in Hz. Two analysis rates have been considered. Firstly, the analysis rate was set to 250 Hz, realized with a 50% overlap between two successive analyses frames. Secondly, a 500 Hz analysis rate was achieved with a 75% overlap. Let us recall that the speech signal was initially sampled at 16 kHz. Each frame contained 128 points and lasted 8 ms. At the 250 Hz analysis rate, the updated stimulus was refreshed every 4 ms, corresponding so to a 50% overlap. At the 500 Hz, the stimulus was updated every 2 ms (75% overlap).

3. Method

In the present study, the influence of several stimulation modes on speech intelligibility in quiet and in noisy environments was investigated. Signal was "vocoded" using the same algorithms used in CI processing. Processed speech signals were presented to normal hearing subjects. To assess the performance of the aforementioned strategies, comprehensive phonemes recognition tests were conducted.

3.1. Listeners

The proposed speech processing algorithms were assessed by a group of fifty normal hearing subjects. Prior to the experiment sessions, their audiogram was tested in the ENT department of the Edouard-Herriot hospital of Lyon. Subject's age ranged from 18 to 32 years. All participants were French native speakers. Tests with all the subjects were carried in an anechoic room.

3.2. Stimuli

The phonetic material was selected from the French Lafon set which contains twenty lists of seventeen 3-phoneme words pronounced by a single male talker [23]. The sound level was calibrated at 70 dB SPL corresponding to a classical comfortable level. The interfering noise was multi-talker speech babble. The distance between the Loud Speakers (LS) and the Kemar (Knowledge Electronic Manikin for Acoustic Research) manikin was about 2 m. Four SNRs were used: NAN (No Added Noise), 6 dB, 0 dB and −6 dB. The target speech signal was placed directly in front of the Kemar manikin at 0° azimuth (LS3). Speech signals were corrupted by additional noise coming from two noise loudspeakers placed symmetrically in both hemi-fields (−90° and 90°: LS1 and

LS5). The experimental setup is presented in Fig. 4. The recorded speech signals were processed using the previously described 'vocoder' algorithm and then presented to the normal hearing subjects. Two lists (34 words, thus 102 phonemes) were presented to the listeners.

3.3. Listening session

The listening tests were conducted using a personal computer connected to a CD player (PHILIPS-CD723). An audiometer (MADSEN-Orbiter 922) was used for calibration and intensity level adjustment. Stimuli were presented to the subjects though closed professional 'Sennheiser' HD250 headphones at both ears. Speech stimuli were processed, prior to the listening sessions, with MATLAB software. Stimuli were generated in noise free and noisy conditions with SNRs of 6 dB, 6 dB and −6 dB.

Prior to the formal testing, a training session containing ten random words was delivered in order to familiarize each subject with the stimuli. No score was calculated in the training session. Following this training session, the subjects were tested in various experimental conditions. During the testing session, the subjects were instructed to repeat the words they heard. In total, there were 16 testing conditions (2 CI modes × 2 analysis rates × 4 SNRs). A sequential test order, starting with words processed in quiet and then in noise from the highest SNR (6 dB) and to the lowest SNR (−6 dB) was employed. This sequential approach was chosen in order to give the subjects some adaptation time before listening in noisy conditions. At the end of each listening session, responses were collected, stored and scored off-line according to the number of correctly identified phonemes. All phonemes were scored. The percentage of correctly repeated phonemes out of 102 (two of Lafon's lists) was then calculated.

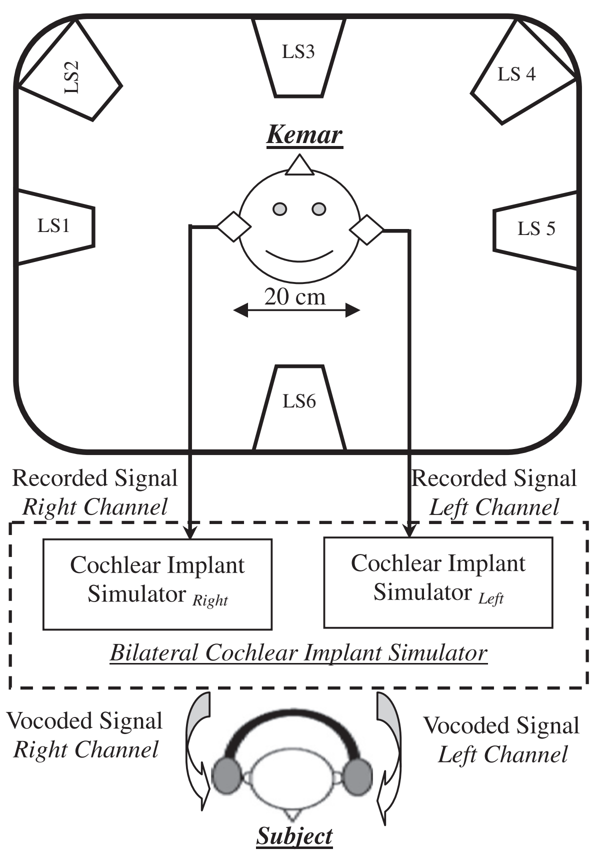

Fig. 4. Experimental setup (Anechoic room) LS3 = speech, LS1 and LS5 = noise. LS2, LS4 and LS6 were not used.

3.4. Results

3.4.1. Signal representation

For a signal processing point of view, a 3D representation of the original speech and processed speech signals via both SBCI and ShBCI CI modes (500 Hz analysis rate) in silent environment can be illustrated in Fig. 5. In fact, the power levels (in dB) for each frame of data and for each frequency channel 'm' ($m = 1 \cdots 12$) are represented. The example is the French word 'agis' (act) taken in the first Lafon's list. The overall 3-D shape of the signal is kept, but "vocoding" introduced some loss at the beginning of the utterance.

3.5. Statistical analysis

Intelligibility scores for this experiment were the percentage of correctly repeated phonemes per test condition. Chi2 analysis with the following parameters was performed to examine the effects of the considered factors:

– Repeated measures (the same subjects underwent all the situations).
– Dependant variable: the percentage of recognition (score).
– Three factors:
• The mode (SBCI, ShBCI).
• The analysis rate (250 Hz and 500 Hz).
• The noise level (this last factor was considered as random; the first 2 factors were fixed).

This model, with both fixed and random effects, is a mixed-effect model and is a standard for data analysis (Baayen, 2008). The 'lmer' program of the 'lme4' package [2] of the statistical programming language and software environment R [28] was used.

Results indicated an effect of the mode (Chi2[1] = 112, $p < 0.001$), the analysis rate (Chi2[1] = 72; $p < 0.001$) and, obviously, the SNR (Chi2[3] = 16,305; $p < 0.001$).

In addition, there was a significant interaction between the mode and the SNR (Chi2[3] = 15; $p < 0.005$). No significant interaction was seen between the analysis rate and the SNR (Chi2[3] = 0.7; $p = 0.8$). No significant interaction appeared between the mode and the analysis rate (Chi2[1] = 2.6; $p = 0.1$). Fig. 6 indicates the mean percent scores on phoneme recognition as a function of the stimulation mode in quiet and in noisy environment at different SNR levels.

The results in Fig. 6 indicates that the performance depended on both CI mode and the analysis rate in quiet and in noisy environments at different SNRs. Also, phoneme recognition score was strongly affected by the SNR. Overall improvement of ShBCI *vs* SBCI was 4%. Chi-2 tests confirmed the existence of the mode effect. Mean recognition percentages are indicated on Table 2.

Post-hoc comparisons were run to explore the differences obtained with the considered stimulation modes at different SNR levels. We used the Tukey HSD test via the general linear hypothesis test 'glht' function from multiple comparison 'multcomp' package of R.

In quiet (No Added Noise: NAN), phoneme recognition was affected by the stimulation mode. Post-hoc comparison indicated that phoneme recognition performance improved using ShBCI over SBCI and the improvement was about 2.5%; but this difference was not significant ($p > 0.005$). When SNR was 6 dB, a significant improvement of phoneme recognition was observed when ShBCI mode was considered ($p < 0.001$). In fact, better phoneme recognition at 6 dB when the ShBCI mode was adopted and the improvement was 4%. When signal and noise levels were identical, a significant improvement of 5.3% for ShBCI over SBCI was observed ($p < 0.005$). A similar pattern was observed at −6 dB SNR. ShBCI

Please cite this article in press as: Kallel F et al. Influence of a shift in frequency distribution and analysis rate on phoneme intelligibility in noisy environments for simulated bilateral cochlear implants. Appl Acoust (2012), http://dx.doi.org/10.1016/j.apacoust.2012.05.014

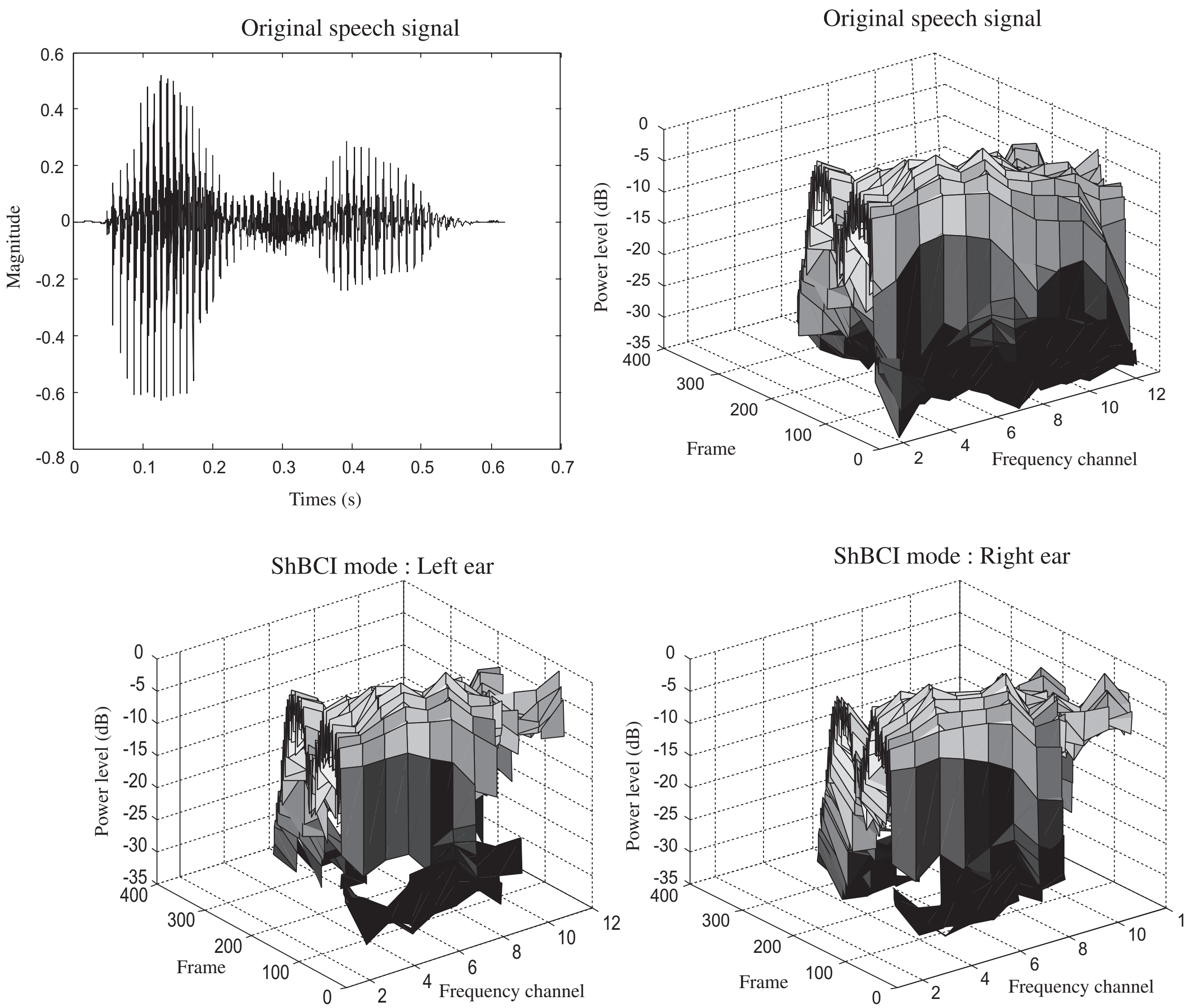

Fig. 5. 3D representation of the French words 'agis'. For the symmetrical stimulation mode (SBCI), both signals are identical to ShBCI left ear.

was significantly better than SBCI ($p < 0.005$). Improvement obtained with ShBCI was less than seen for 0 dB (4.1%).

4. Discussion

4.1. Shifted vs symmetrical (ShBCI vs SBCI)

Two modes of bilateral CI were considered in the current study: SBCI and ShBCI. Better results were seen with ShBCI mode, especially in noisy conditions. In the case of ShBCI, the frequency bands were not the same on both ears as there was a frequency shift between the two ears. This frequency shift, driven by the bark scale, was delivered to the auditory system. It increases the redundancy in the speech signal and subsequently can improve signal intelligibility.

Now, it has to be seen with CI recipients if ShBCI mode (shifted) is more beneficial than SBCI (symmetrical). It is well known that, with implantees, each ear presents differences in neural survival increasing the pattern differences transmitted to the brain. CI stimulation indicated a difference and a shifted mode may amplify this effect.

4.2. Analysis rate

The refreshing rate was the second parameter considered in this study. Low analysis rate (250 Hz) and a moderate analysis rate (500 Hz) were examined. Rate can go up to 1000 Hz in the ACE strategy. An average improvement of 3.3% in favor of the moderate analysis rate (500 Hz) was seen. In quiet, the analysis rate presented no significant effect (difference was 1.5%). When the noise was added, difference was more important and 500 ups was better than 250 ups (2.6% at 6 dB, 5.1% at 0 dB and 3.8% at −6 dB).

4.3. Consistency with the literature

Current work findings are consistent with previous studies done with Nucleus devices using the ACE strategy [36,14,24,39]. In these studies, no effect of the stimulation rate for monosyllabic word or consonant perception was seen. In this current work, results were rather similar in quiet, but a difference appeared when noise was added. This study agrees with speech perception outcomes obtained in quiet.

In Weber et al. [39], the authors studied the performance obtained with ACE strategy at 500, 1200 and 3500 pps; results did

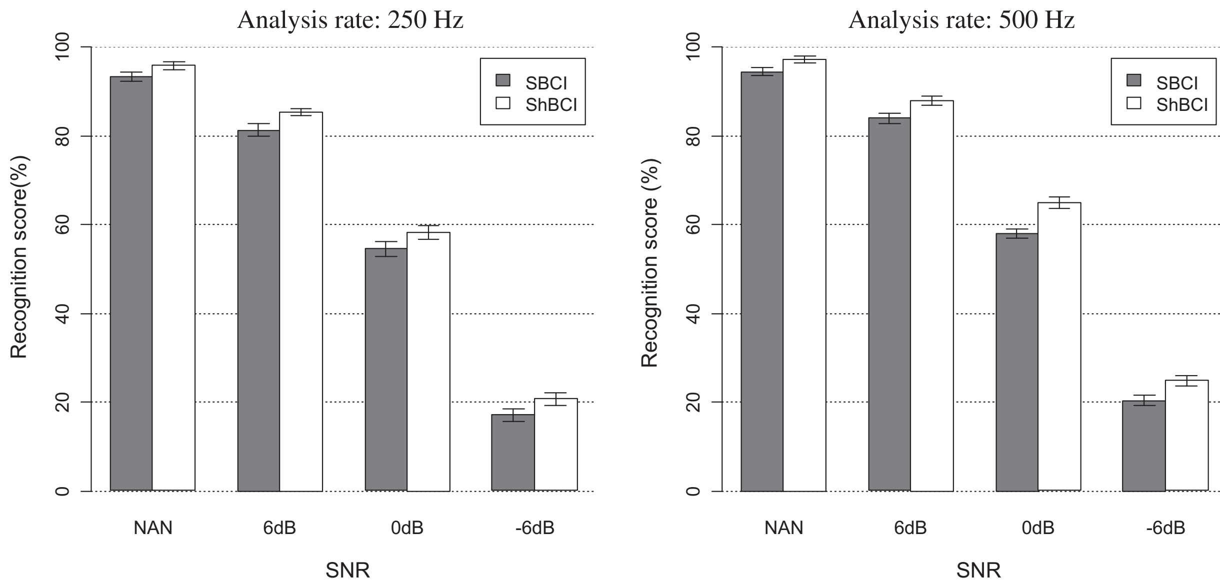

Fig. 6. Recognition score as a function of stimulation mode for different SNR levels at two different analysis rates.

not show a difference for the ACE strategy for rates ranging from 500 to 3500 pps. An evaluation of the effects of stimulation rate on performance with the Nucleus Freedom CI was presented in Balkany et al. [1]. Results pointed out that 67% of implantees preferred slow stimulation rates and the performance was not improved with higher stimulation rates. Going further, results in this current work are consistent with clinical trials done by Komal et al. [15]; the authors presented a study on eight postlingually deaf adults, users of the Nucleus CI24, and they explored speech perception in quiet and in noise for low to moderate stimulation rates. Their results, for six out of the eight tested subjects, showed no significant effect of the stimulation rate for monosyllables recognition in quiet. But, results for the syllable test in noise indicated improvements with 500 or 900 pps, compared to 275 and 350 pps (for seven out of the eight subjects).

5. Conclusion

The current work investigated phoneme recognition through an acoustic CI simulator built according to the Neurelc Digisonic SP CI parameters. Two CI modes and two analysis rates have been considered. Phoneme recognition performances were measured with and without a competing noise source at different SNR levels. Result indicated a significant interaction between CI mode and SNR. It was seen that a shifted bilateral stimulation mode (ShBCI) was more efficient than the symmetrical bilateral stimulation mode (SBCI). An improvement in speech intelligibility was found with ShBCI and it is thought to be due to some redundancy introduced by the frequency shift and positively interpreted by the brain. The envelope update rate was also varied to simulate different analysis speeds. 250 Hz and 500 Hz updating rates were evaluated and improvements were found with higher update rate, in the presence of noise.

Acknowledgments

The authors are grateful to several people for their support in this study: Dr. Mohamed Ghorbel and Pr. Hung Thai Van for their suggestions during the course of the work, Dr. Stephane Gallego for his assistance in building up the CI coding strategy and Pr. Lionel Collet for the facilities he provided. We are also grateful to the Neurelec Company and to Dr. Dan Gnansia for the assistance in upgrading the manuscript.

This study was funded by Egide (code: 09G 1421) and CNRS-DGRST (code: 09/R11-18) Projects for French–Tunisian collaboration.

References

[1] Balkany T, Hodges A, Menapace C, Hazard L, Driscoll C. Nucleus freedom North American clinical trial. Otolaryngol Head Neck Surg 2007;136:757–62.
[2] Bates DM. Linear mixed model implementation in lme4. Manuscript, university of Wisconsin-Madison; 2007.
[3] Carroll J, Zeng FG. Fundamental frequency discrimination and speech perception in noisecochlear implant simulations. Hear Res 2007;231(1–2):42–53.
[4] Cooper HR, Craddock LC. Cochlear implants-a practical guide. 2nd ed. West Sussex, England: Whurr Publishers Limited; 2006. p. 1–69.
[5] Chouard CH, Weber B, Chabolle F, Fugain C. Les implants cochléaires monocanal monomac et multicanaux minimac. Annale Otolaryngol 1988;105:227–36.
[6] Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble1 W. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. Ear Hearing 2010;29:352–9.
[7] Dunn CC, Tyler RS, Oakley S, Gantz BJ, Noble1 W. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. Ear Hearing 2008;29:352–9.
[8] Drennan WR, Won JH, Dasika VK, Rubinstein JT. Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise. J Assoc Res Otolaryngol 2007;8:373–83.
[9] Dillon H. Hearing aids. Turramurra (N.S.W): Boomerang Press; 2001.
[10] Dudley H. The vocoder. Bell Labs Record 1939;17:122.
[11] Dorman MF, Loizou PC, Fitzke J, Tu Z. The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. J Acoust Soc Am 1998;104:3583–5.
[12] Fourakis M, Hawks J, Holden L, Skinner M, Holden T. Effect of frequency boundary ACE speech coding strategy. J Am Acad Audiol 2004;15:281–9.
[13] Helms J, Müller J, Schön F, Winkler F, Moser L, Shehata-Dieler W. Comparison of the TEMPO+ ear-level speech processor and the CIS PRO+ body-worn processor in adult MED-EL cochlear implant users. J Oto-Rhino-Laryngol Relat Spec 2001;63:31–40.
[14] Holden LK, Skinner MW, Holden TA, Demorest ME. Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. Ear Hearing 2002;23:463–76.
[15] Komal A, Pam D, Richard D, Andrew V. Electrical stimulation rate effects on speech perception in cochlear implants. Int J Audiol 2009;48:561–7.
[16] Kasturi K, Loizou PC. Effect of filter spacing on melody recognition: acoustic and electric hearing. J Acoust Soc Am 2007;122(2):EL29–34.
[17] Loizou PC, Hu Y, Litovsky R, Yu G, Peters R, Lake J, et al. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. J Acoust Soc Am 2009;125:372–83.

[18] Loizou PC. Speech processing in vocoder-centric cochlear implants. Adv Otorhinolaryngol 2006;64:109–43.
[19] Loizou PC, Poroy O, Dorman MF. The effect of parametric variations of cochlear implant processors on speech understanding. J Acoust Soc Am 2000;108:790–802.
[20] Loizou PC. Mimicking the human ear: an overview of signal processing techniques for converting sound to electrical signals in cochlear implants. IEEE Signal Process Mag 1998;15:101–30.
[21] Laneaua J, Moonen M, Wouters J. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. J Acoust Soc Am 2006;119:491–506.
[22] Laszig R, Aschendorff A, Stecker M. Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. Otol Neuro-otol 2004;25:958–68.
[23] Lafon JC. In: Eindhoven, editor. Le test phonétique et la mesure de l'audition. Centrex; 1964.
[24] Plant K, Holden L, Skinner M, Arcaroli J, Whitford L. Clinical evaluation of higher stimulation rates in the Nucleus research platform 8 system. Ear Hearing 2007;28:381–93.
[25] Parkinson AJ, Arcaroli J, Staller SJ, Arndt PL, Cosgriff A, Ebinger K. The Nucleus 24 contour cochlear implant system: adult clinical trial results. Ear Hearing 2002:41S–8S.
[26] Peissig J, Kollmeier B. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J Acoust Soc Am 1997;101:1660–70.
[27] Ricketts TA, Grantham DW, Ashmead HD, Haynes DH, Labadie RF. Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources. Ear Hearing 2006;27:763–73.
[28] R development core team. R: A language and environment for statistical computing. Vienna: R; 2007.
[29] Rubinstein JT, Parkinson WS, Tyler RS, Gantz BJ. Residual speech recognition and cochlear implant performance. effects of implantation criteria. Am J Otol 1999;20:445–52.
[30] Schleich P, Nopp P, D'Haese P. Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40 cochlear implant. Ear Hearing 2004;25:197–204.
[31] Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–4.
[32] Tranmüller H. Analytical expression for the tonotopic sensory scale. J Acoust Soc Am 1990;85:97–100.
[33] Veekmans K, Ressel L, Mueller J, Vischer M, Brockmeier SJ. Comparison of music perception in bilateral and unilateral cochlear implant users and normal-hearing subjects. Audiol Neurotol 2009;14:315–26.
[34] Verschuur CA. Effect of stimulation rate on speech perception in adult users of the Med-El CIS speech processing strategy. Int J Audiol 2005;44:58–63.
[35] Van Hoesel RJ. Exploring the benefits of bilateral cochlear implants. Audiol Neurotol 2004;9:234–46.
[36] Vandali AE, Whitford LA, Plant KL, Clark GM. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. Ear Hearing 2000;21:608–24.
[37] Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J Assoc Res Otolaryngol 2007;8:384–92.
[38] Wilson B, Finley C, Lawson D, Wolford R, Eddington D, Rabinowitz W. Better speech recognition with cochlear implants. Nature 1991;352:236–8.
[39] Weber BP, Lai WK, Dillier N, von Wallenberg EL, Killian MJ. Performance and preference for ACE stimulation rates obtained with Nucleus RP 8 and freedom system. Ear Hearing 2007;28:465–85.
[40] Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. J Acoust Soc Am 2007;122:2376–88.
[41] Zeng FG, Fay RR. Cochlear implants: auditory prostheses and electric hearing. New York: Springer; 2004. p. 334–376.
[42] Zwicker E, Fast H. Psychoacoustics facts and models. Berlin: Springer Verlag; 1990.