

Procédé de correction auditive utilisant un modèle paramétrique du signal de parole¹

Philippe GOURNAY¹, Gilles QUAGLIARO¹, Alain GOYE², Gwénaél GUILMIN¹, Frédéric CHARTIER¹

¹Thomson-CSF Communications
66, rue du fossé blanc - BP 156 - 92231 GENNEVILLIERS CEDEX FRANCE

²Laboratoire CRISTAL
Service ORL - Hôpital Avicenne - 125, rue de Stalingrad - 93009 Bobigny CEDEX FRANCE

speech@tcc.thomson-csf.com, alain.goye@avc.ap-hop-paris.fr

Résumé – Cet article présente un nouveau procédé de correction auditive utilisant un modèle paramétrique du signal de parole. Ce procédé permet, en plus des opérations classiques de compression et d’amplification, d’effectuer des transformations élaborées (modification du rythme temporel, modification de l’enveloppe spectrale, etc.) tout en conservant la structure naturelle du signal de parole. Les premiers résultats d’une série de tests cliniques montrent que ce procédé est bien accueilli par les personnes atteintes de surdités intermédiaires (moyennes ou sévères) et pour lesquelles il n’existe pas véritablement de traitement prothétique adapté.

Abstract – This paper presents a new signal processing technique for hearing aids that makes use of a parametric model of the speech signal. This technique allows, in addition to the classical compression and amplification operations, to implement some elaborate transformations (modification of the temporal rhythm, modification of the spectral envelope, etc.) while preserving the natural structure of the speech signal. The first results of a series of clinical evaluations show that this technique is beneficial to people with intermediate (middle or severe) hearing impairment, for whom no really satisfying auditory compensation technique is available.

1. Introduction

Les surdités sont traditionnellement classées en quatre catégories. Les surdités les moins handicapantes sont dites « légères » : elles sont efficacement corrigées par les aides auditives conventionnelles (prothèses, contours d’oreille), qui mettent en œuvre des opérations d’amplification et de compression dans un nombre réduit de sous-bandes (typiquement 3). Les surdités les plus handicapantes sont dites « profondes ». Elles peuvent parfois être corrigées par la pose d’un implant cochléaire, qui agit par stimulation directe du nerf auditif au moyen d’électrodes [1]. Pour les surdités intermédiaires, dites « sévères » ou « moyennes », les aides auditives conventionnelles sont souvent inefficaces (pas de correction possible pour les plages de fréquence dans lesquelles il ne subsiste pas d’audition, risque d’effet Larsen par sur-amplification, etc.), et le risque lié à la pose d’implants cochléaires est important (perte définitive du reliquat d’audition).

Un certain nombre de procédés ont été étudiés pour redonner une certaine intelligibilité au signal de parole perçu par les sourds profonds. Ces traitements exploitent le fait qu’il subsiste souvent chez ces personnes un reliquat de perception

(généralement dans les graves) dont il est possible de tirer parti. La référence [2] par le professeur J.C. Lafon présente un bon état de l’art dans le domaine. La technique imaginée en 1952 par J.M. Tato consiste à enregistrer un signal de parole prononcée très rapidement et à le restituer à vitesse moitié. Des essais ont montré que cette opération, qui réalise une transposition de la parole d’une octave plus bas tout en conservant sa structure naturelle, présente un certain intérêt pour les sourds ; elle ne se prête malheureusement pas à un traitement en temps réel. En 1971, MM. C. Vigneron et M. Lamotte ont tenté d’adapter cette méthode au temps réel, en découpant la parole en intervalles de 1/100 de secondes, en supprimant un intervalle sur deux, et en appliquant le procédé de J.M. Tato sur les intervalles restant. La parole produite par ce procédé présentait toutefois un bruit trop important. Enfin, la prothèse GALAXIE proposée par J.C. Lafon met en œuvre une batterie de filtres et de mélangeurs répartis sur six sous-bandes, et réalise une transposition dans les plages de fréquence perceptibles par l’utilisateur. Ce type de procédé, parce qu’il ne conserve pas la structure naturelle de la parole, n’a pas été jugé satisfaisant par les malentendants.

¹ Ce travail a été subventionné par l’AGEFIPH (Association nationale pour la GEstion du Fonds pour l’Insertion Professionnelle des personnes Handicapés) et HANDI-THOMSON.

Pour conclure ce bref état de l'art, on peut dire que - même s'il existe un certain recouvrement entre les domaines de prescription de l'aide auditive et de l'implant cochléaire - il n'existe pas vraiment d'offre prothétique adaptée à la catégorie des surdités intermédiaires. Toutefois, il apparaît clairement à la lecture de la référence [2] que :

1. Il est important de pouvoir transposer la globalité de la structure acoustique de la parole, c'est à dire de ramener ses éléments structurels porteurs d'intelligibilité dans la zone de perception du malentendant.

2. Il est également important de produire des sons « naturels », c'est à dire de produire une parole modifiée dont la structure reste en harmonie avec l'acquis auditif du malentendant.

3. Enfin, il faut veiller à conserver la structure temporelle du signal car le rythme est porteur d'information accessible au malentendant.

Cet article présente un nouveau procédé de correction auditive, pour les surdités sévères et moyennes, utilisant un modèle paramétrique du signal de parole. Ce procédé permet, en plus des traitements « conventionnels » d'amplification et de compression sélective en fréquence, de réaliser un certain nombre de traitements « spéciaux » (compressions énergétiques, modifications de l'enveloppe spectrale, ralentissement temporel, etc.) susceptibles de ramener une partie de l'intelligibilité du signal de parole dans la plage d'audition du malentendant tout en conservant la structure naturelle du signal de parole. Quelques résultats préliminaires d'une campagne de tests cliniques sont donnés.

2. La modélisation paramétrique du signal de parole

Les procédés de traitement numérique du signal appliqués aux signaux de parole se basent souvent sur une modélisation paramétrique du signal de parole [3]. Ce modèle comporte généralement quatre paramètres (voir figure 1) :

- Une information de « voisement », qui décrit le caractère plus ou moins périodique (sons voisés) ou aléatoire (sons non voisés) du signal ;
- La fréquence fondamentale ou « PITCH », pour les sons voisés ;
- L'évolution temporelle de « l'énergie » du signal de parole ;
- Son enveloppe spectrale, ou « spectre », qui est généralement obtenue par une modélisation auto-régressive (filtre de prédiction linéaire ou LPC) ou par une analyse de Fourier à court terme synchrone avec le pitch.

Ces quatre paramètres sont estimés périodiquement sur le signal de parole (de une à plusieurs fois par trame selon le paramètre, pour une durée trame typiquement comprise entre 10 et 30 ms).

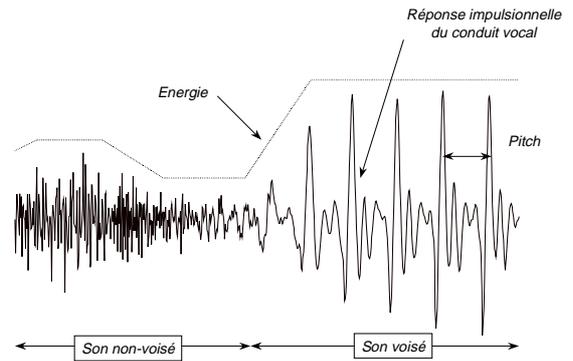


FIG. 1 : paramètres de modélisation du signal de parole

Dans le domaine des communications numériques, il existe de nombreux contextes dans lesquels la capacité du canal de transmission est très fortement limitée (téléphonie par satellite, télécommunications dans la gamme HF, etc.). Il est alors souvent fait appel à cette modélisation pour diminuer le débit numérique nécessaire à la transmission de la parole : c'est le principe du vocodeur (figure 2). Au niveau de l'émetteur, une procédure d'analyse réalise l'estimation des paramètres du modèle. Une procédure de compression (souvent appelée « codage ») permet de réduire le débit binaire associé au signal de parole en vue de sa transmission par un système numérique. L'exemple le plus simple de codage est celui de la quantification. Au niveau du récepteur, une procédure de synthèse assure la reconstitution d'un signal de parole à partir des paramètres du modèle.

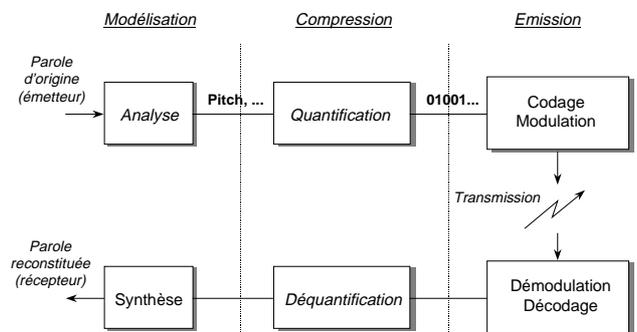


FIG. 2 : système de transmission numérique de la parole

Ainsi, un vocodeur à 1200 bits par seconde permet de transmettre plus de 95 % de l'intelligibilité d'un signal de parole [4], ce qui représente un facteur de compression supérieur à 50 par rapport à de la téléphonie numérique classique (standard MIC à 64 kbits/s).

Dans le cas d'une surdité, le récepteur n'est pas une personne distante mais une personne dont la plage de perception auditive, en bande passante ou en dynamique, est limitée. L'objectif du nouveau procédé de correction auditive est donc de réaliser une compression du signal de parole, non pas au sens de la quantité d'information binaire à transmettre, mais au sens de la bande de perception nécessaire à sa compréhension.

3. Description du procédé

3.1 Principe général

Le procédé de traitement est appliqué entre le locuteur « naturel » et l'oreille malentendante (figure 3) ; il vise à transposer l'intelligibilité de la parole d'origine dans la zone de perception du malentendant tout en lui conservant sa structure naturelle.

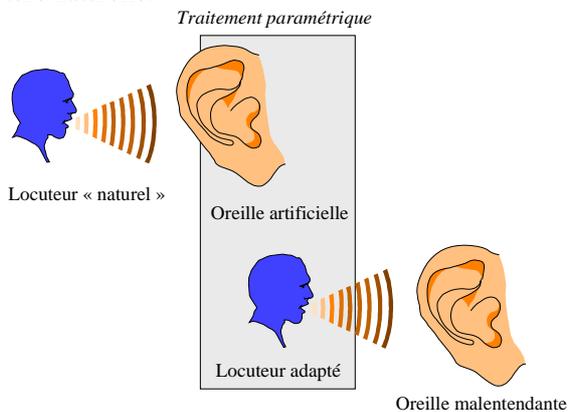


FIG. 3 : principe de base du procédé de correction auditive

3.2 Le procédé de correction auditive

Le procédé (figure 4) consiste tout d'abord en un certain nombre de pré-traitements, parmi lesquels on peut citer :

- La réduction du bruit de fond ;
- L'annulation d'écho acoustique ;
- Le contrôle manuel ou automatique de gain.

Ces pré-traitements sont suivis par une analyse paramétrique du signal de parole, la modification des paramètres du modèle, et la synthèse d'un signal de parole « modifié ». Des opérations de :

- compression énergétique,
- modification du contenu spectral (pitch et voisement),
- modification de l'enveloppe spectrale,
- et de modification du rythme temporel

utilisant le modèle paramétrique et adaptées aux caractéristiques de l'auditeur sont mises en œuvre. Les procédures d'analyse et de synthèse de la parole sont basées sur la modélisation HSX [4]. Le problème de la sensibilité au bruit de fond des traitements paramétriques est bien connu ; la référence [5] démontre l'intérêt d'utiliser une procédure de réduction du bruit de fond acoustique. La modification d'enveloppe spectrale peut être une compression homothétique, qui n'est pas sans rappeler celle utilisée pour corriger la parole hyperbare (cet effet « Donald Duck » est rencontré par les plongeurs en eaux profondes qui respirent un mélange de gaz légers – hélium, hydrogène – en plus de l'oxygène. Du fait de la faible densité du mélange gazeux remplissant le conduit vocal, l'enveloppe spectrale de la parole subit alors en première approximation une dilatation sur son axe fréquentiel [6]. La parole n'est alors pas compréhensible sans un très long apprentissage).

Ces traitements, qualifiés de « spéciaux », sont suivis par un certain nombre de post-traitements, incluant une égalisation linéaire qui vise à corriger l'audiogramme en amplifiant ou atténuant certaines bandes de fréquences. Enfin, pour une utilisation binaurale, un multiplexage permet de choisir entre une restitution monophonique (par exemple, voie « traitée » seule) et une restitution stéréophonique (par exemple, voie « traitée » d'un côté et « non traitée » de l'autre). Cette dernière permet au malentendant d'adapter le traitement à chacune de ses oreilles (deux égaliseurs pour compenser deux audiogrammes différents), et éventuellement de conserver intact sur une oreille un signal auquel il est habitué et sur lequel il peut s'appuyer pour se synchroniser.

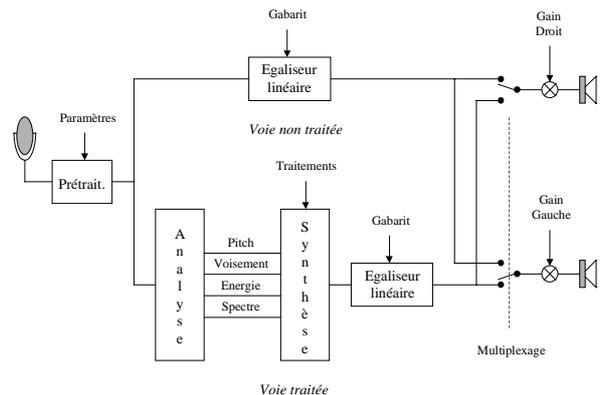


FIG. 4 : schéma bloc du procédé de correction auditive

3.2 Mise en œuvre

Ce procédé a fait l'objet d'une implémentation logicielle sur PC multimédia sous Windows®. Cet « Atelier Parole » (figure 5) permet la configuration des traitements et la gestion de profils personnalisés. Il permet de traiter des fichiers préalablement enregistrés, et de traiter en temps réel les signaux provenant de sources internes (sons système du type « message d'erreur », logiciels d'apprentissage de langue, etc.) ou externes (radio, téléphone, lorsque l'interface matérielle existe). Enfin, un ensemble de représentations graphiques permet de visualiser l'évolution temporelle des paramètres du modèle, ce qui en fait un véritable outil de recherche utilisé lors de la campagne de tests cliniques.



FIG. 5 : l'interface Windows de l'Atelier Parole

4. Évaluation clinique

4.1 Le protocole de test

Compte tenu des multiples possibilités offertes par l'outil, une première série d'essais nous a été nécessaire pour cadrer les transformations a priori les plus intéressantes. Ces premiers tests étaient orientés vers l'évaluation d'une amélioration globale du confort et de l'intelligibilité de la parole bruitée, à la suite de l'une des transformations suivantes : ralentissement dans un facteur 1.5, abaissement du pitch d'un facteur 2, augmentation de la fréquence de voisement d'un facteur 4, transposition de l'enveloppe spectrale d'un facteur 2 après préaccentuation et abaissement du pitch. Cette première série a mis en évidence certaines contraintes d'utilisation du système : sensibilité à un fort niveau de bruit (en l'absence du procédé de réduction de bruit, qui a été mis en place ultérieurement et devrait réduire cette contrainte), difficulté de réglage individuel des traitements « classiques » (égalisation et compression), et biais dû au manque d'acclimatation du sujet au traitement. La deuxième série d'essais devait donc utiliser des conditions où le bénéfice attendu soit immédiatement mesurable, et évaluer l'amélioration du confort auditif autant que de l'intelligibilité.

Nous avons utilisé pour cela un test de discrimination de type VCV, et un test subjectif de confort. Dans le premier, on présente au sujet successivement, dans un ordre aléatoire, 64 logatomes comprenant chacun une consonne parmi 16 dans un contexte de voyelles 'a', avec la consigne de reconnaître la consonne. Ce test permet d'établir une matrice de confusions consonantiques. La comparaison des matrices moyennes permet d'évaluer tout changement dans la capacité de discrimination d'une consonne donnée. Dans le test subjectif de confort, le sujet exprime une préférence entre deux versions, traitée et non traitée, de 5 phrases différentes. Les deux tests étaient répétés de manière à évaluer trois traitements : égalisation, ralentissement et réduction de bruit. Vingt sujets malentendants adultes, sans contrainte sur le profil de perte auditive, ont été soumis au protocole.

4.2 Résultats préliminaires

Les résultats aux tests objectifs suggèrent qu'il existe une catégorie non négligeable de sujets pour lesquels soit une, soit plusieurs des transformations proposées apportent une amélioration de la capacité de discrimination des logatomes utilisés. On n'observe cependant pas d'amélioration significative sur l'ensemble des sujets, ni sur l'une des catégories suivantes : sujets âgés de moins - ou plus - de 60 ans, sujets d'audiogramme relativement plat (écart moyen de moins de 10,2 dB) ou variable. Il est plus probable que les résultats positifs soient corrélés avec un type de perte auditive particulier, éventuellement une forme d'audiogramme. Cependant le petit nombre de patients testés jusqu'ici, au regard de la grande variabilité des profils auditifs, rend difficile d'établir de telles corrélations.

Les tests subjectifs montrent nettement que certains patients préfèrent la forme traitée des stimuli. On note un écart de pourcentage statistiquement significatif ($p < 0,01$) sur l'ensemble des sujets dans le cas du débruitage. Dans certains cas le sujet n'était pas en mesure d'exprimer une préférence ; l'emploi d'un plus grand nombre de phrases aurait aidé à lever cette indétermination. Une catégorie de sujets présente une amélioration du confort pour chacune des transformations proposées ; si cela traduit une acclimatation plus rapide, cela suggère qu'avec le temps les préférences pour les stimuli traités doivent croître. Dans tous les cas, il nous semble positif de noter que sur les 20 sujets testés, trois expriment une préférence pour les trois transformations, neuf pour au moins deux, et quinze pour au moins une d'entre elles. Comme pour les scores de discrimination et d'intelligibilité, ces préférences pourraient probablement être accentuées par un réglage personnalisé des transformations.

5. Conclusion et perspectives

Un nouveau procédé de correction auditive utilisant un modèle paramétrique du signal de parole a été présenté. Ce procédé permet d'effectuer des transformations élaborées (modification du rythme temporel, modification de l'enveloppe spectrale, etc.) tout en conservant la structure naturelle du signal de parole.

Les résultats préliminaires obtenus par une série de tests cliniques montrent que les traitements proposés sont plutôt bien accueillis par les malentendants (apports en termes de confort d'écoute et d'intelligibilité). Les deux principales difficultés rencontrées sont le réglage des nombreux paramètres de traitement en fonction des caractéristiques d'audition, et la nécessité d'un apprentissage (notamment pour la compression homothétique de l'enveloppe spectrale).

Références

- [1] L. Plévert. L'électronique sait aujourd'hui se connecter au corps humain. *Électronique Internationale Hebdo*, n° 294, 12 Février 1998, pages 20-21
- [2] J.-C. Lafon. Transposition et modulation. *Bulletin d'audiophonologie*. Annales scientifiques de Franche Comté, volume XII, n° 3 et 4, monographie 164, 1996
- [3] Calliope. *La parole et son traitement automatique*. Collection Technique et Scientifique des Télécommunications, Masson 1989.
- [4] P. Gournay, F. Chartier. A 1200 bits/s speech coder for very low bit rate communications. *IEEE Workshop on Signal Processing Systems (SiPS'98)*, Boston, 8-10 Octobre 1998.
- [5] G. Guilmin, R. Le Bouquin-Jeannès, et P. Gournay. Study of the influence of noise pre-processing on the performance of a low bit rate parametric speech coder. *Eurospeech'99* (a paraître).
- [6] J. Crestel / M. Guitton. *Contribution à l'amélioration de l'intelligibilité de la communication orale en milieu hyperbare*. Thèses de Doctorat de l'Université de Rennes, 1987