

Optimisations d'un correcteur de parole hyperbare

Zahir Smail et Jean Claude Malherbe

Groupe de REcherche en Science de l'Ingénieur de Lannion
I.U.T. de Lannion
BP 150
22300 Lannion

RÉSUMÉ

Cet article traite du problème de la correction de la parole hyperbare, caractéristique des plongeurs professionnels, par la méthode du vocodeur à canaux numérique modifié. Une nouvelle approche de correction est proposée, elle permet d'éviter les problèmes liés à la loi de correction utilisée, limitant le fonctionnement des correcteurs usuels à des profondeurs n'excédant pas 300m. Le correcteur ici soumis est également modifié afin de réduire le temps de traitement. Les optimisations apportées ont permis d'une part d'éviter la limitation en profondeur et d'autre part de réduire le temps de traitement de l'ordre de 60 %. Des tests d'écoute subjectifs ont montré que la parole corrigée par le nouveau système est proche de la parole naturelle.

ABSTRACT

This paper deals with a technique for restoring the spectral properties of the speech of deep-sea divers operating under various conditions of gas mixtures and diving depths. The method used here is the modified digital channel vocoder. Improvement is first obtained by inverting the previously-used algorithm. It enables correction to be made at any depth, even below the previous 300 metre limit's, which is of great interest. Multirate sampling considerably reduces the processing time, hence improves real-time operation to the order of 60%. Subjective listening tests have shown that speech processed by the new method is close to natural speech.

1 Introduction

Pour résoudre les problèmes physiologiques liés à l'air pressurisé, la plongée profonde utilise un mélange respiratoire à base d'hélium (héliox, hydrox,...). L'inhalation de ce mélange hyperbare altère l'intelligibilité de la parole. Cette déformation de la parole, souvent qualifiée "d'effet Donald Duck" est due, d'une part, à la modification de la masse volumique du gaz inhalé et, d'autre part, à la variation de la vitesse de propagation du son par rapport à l'air à la pression atmosphérique.

Au niveau du signal de parole, la modification de ces deux paramètres génère un déplacement des formants vers les hautes fréquences. On observe par ailleurs une influence sur la valeur de la fréquence fondamentale [6]. Le déplacement des formants, principalement responsable de l'inintelligibilité, est modélisé par la loi (1) de Fant-Linqvist [3], validée par différents chercheurs [12].

$$f_h^2 = k_0^2 f_a^2 + k_0^2 (m-1) f_0^2 \quad (1)$$

où :

f_a et f_h : sont les fréquences d'un même formant respectivement dans l'air et dans le mélange hyperbare

f_0 : est la fréquence de résonance du conduit vocal fermé aux lèvres (air, 1atm) [7].

et avec :

$$k_0 = \frac{ch}{c_a} = \frac{\text{(vitesse du son dans le mélange hyperbare)}}{\text{(vitesse du son dans l'air et à la pression atmosphérique)}}$$

$$m = \frac{\rho_h}{\rho_a} = \frac{\text{(masse volumique du mélange hyperbare)}}{\text{(masse volumique de l'air à la pression atmosphérique)}}$$

L'évolution de ces deux paramètres en fonction de la pression, principale cause de la distorsion de cette parole, est représentée sur les figures 1 et 2 pour les deux mélanges respiratoires « héliox » et « hydrox » [11].

La composante linéaire $k_0 f_a$ de la relation (1) est liée à la variation de la vitesse de propagation du son dans le milieu hyperbare comparée à son équivalent obtenu dans les conditions normales (air à la pression atmosphérique). La composante non linéaire $k_0^2 (m-1) f_0^2$ qui affecte principalement les faibles fréquences est due aux vibrations des parois du conduit vocal [4]. Une analyse plus fine de l'enveloppe spectrale met également en évidence une perte d'énergie des formants et une variation de leur bande passante. Ces deux paramètres n'évoluent pas de façon linéaire sur tout le spectre. La méthode de correction à mettre en œuvre doit tenir compte des modifications de ces différents paramètres pour être optimale.

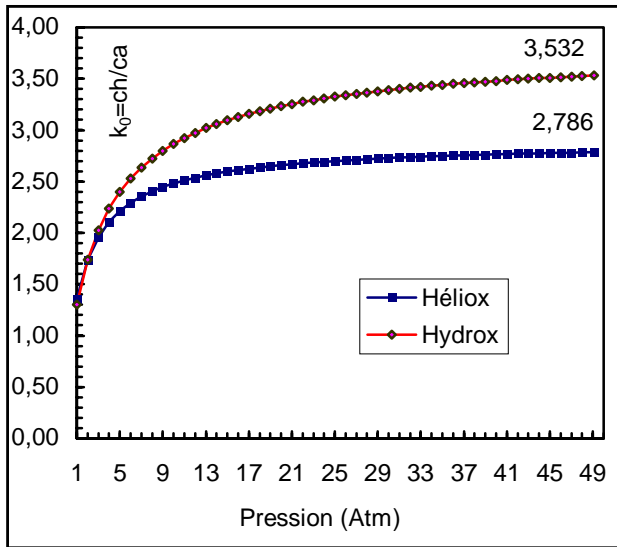


Figure 1 : Evolution du paramètre « k » en fonction de la pression.

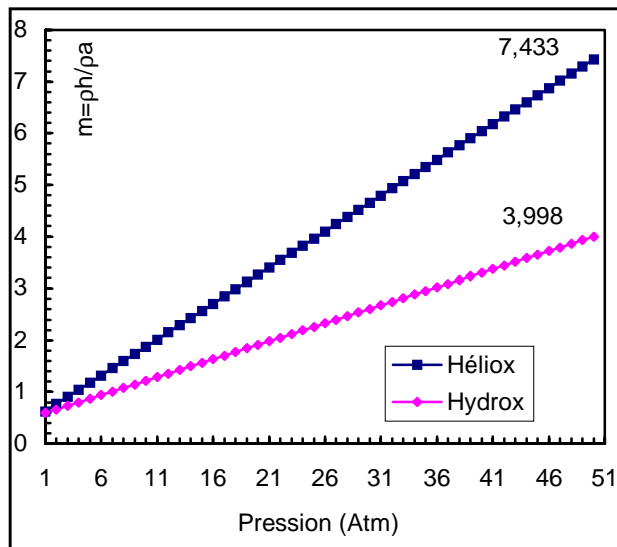


Figure 2 : Evolution du paramètre « m » en fonction de la pression.

2 méthodologie

2.1 Le transcodeur hyperbare.

L'outil de correction ici proposé est celui du vocodeur à canaux numérique (figure 3) à structure modifiée [10] dont le but est d'optimiser la restitution de la fréquence fondamentale.

Suivant son architecture usuelle, une batterie de N filtres passe bande contigus (fréquence centrale f_{hi} , bande passante B_{ai} , $1 \leq i \leq N$) analyse le signal hyperbare et restitue son enveloppe spectrale court terme après un redressement et un filtrage passe bas [2]. A la sortie de chaque canal d'analyse

on obtient l'amplitude moyenne A_{mi} du signal correspondant à la fréquence f_{hi} .

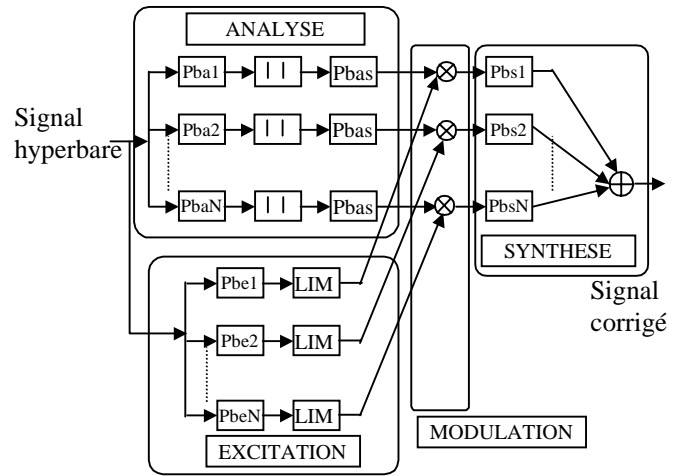


Figure 3 : Structure du correcteur de la parole hyperbare.

Avec :

- Pbai : filtre passe bande d'analyse ($i=1, \dots, N$) ;
- Pbei : filtre passe bande d'excitation ($i=1, \dots, N$) ;
- Pbsi : filtre passe bande de synthèse ($i=1, \dots, N$) ;
- Pbas : filtre passe bas ;
- LIM : limiteur ;
- || : redressement.

Un ensemble de N filtres d'excitation paramétrés par leur fréquence centrale f_{ai} et leur bande passante B_{ai} , qui sont suivis de limiteurs dont la sortie est -1 ou +1 selon le signe de l'entrée, permet de restituer la structure fine de l'excitation. Ces signaux d'excitation sont en suite modulés en amplitude par les intensités A_{mi} de même indice estimées par les canaux d'analyse. Les fréquences f_{ai} correspondent aux fréquences f_{hi} corrigées, leurs valeurs sont déterminées par la relation (2) déduite de la loi de Fant-Linqvist :

$$f_{ai} = \sqrt{\frac{f_{hi}^2}{k_0^2} - (m-1) f_0^2} \quad (2)$$

Les signaux modulés, caractérisés désormais par une intensité A_{mi} et une fréquence corrigée f_{ai} , sont tout d'abord filtrés par les filtres du banc de synthèse (identiques aux filtres d'excitation) pour éliminer les harmoniques supérieurs causés par les limiteurs et sont ensuite sommés pour synthétiser le signal final corrigé. Les filtres du correcteur usuel sont des FIR [9].

Cependant l'observation de la relation (2) montre qu'à partir d'une certaine profondeur (ou k_0 élevé) la quantité sous le radical devient négative, ce qui pose évidemment un problème. Ce phénomène concerne les fréquences de coupures basses (f_{hi} faibles) et les grandes profondeurs (k_0 élevé). Il apparaît vers 300 m pour l'algorithme mis en place par Saadane en 1989 [8].

Par ailleurs la qualité de la parole corrigée dépend étroitement des largeurs de bande des filtres de synthèse [5]. Si ces bandes sont trop étroites ou trop larges des distorsions de l'enveloppe spectrale apparaissent. Ces dernières peuvent affecter la qualité de la parole synthétisée. Pour les sons voisés, si ces largeurs sont trop étroites, la distorsion proviendra de l'absence des harmoniques de la fréquence fondamentale dans certains canaux d'excitation. Cette absence se traduit par la modulation de l'énergie estimée à la sortie des canaux d'analyse par du bruit à la place de l'harmonique souhaité. De même si ces bandes sont trop larges, la distorsion sera causée par la transmission de plus de deux harmoniques par canal d'excitation. Avec la technique de correction décrite ces largeurs sont imposées par celles du banc d'analyse et elles sont en général trop étroites pour les premiers filtres.

Les deux inconvénients de la méthode décrite vont recevoir une solution dans la nouvelle approche proposée.

2.2 Nouvelle approche de correction.

Notre approche consiste à réaliser la compression de l'enveloppe spectrale de manière indirecte. On fixe cette fois en premier les caractéristiques du banc de synthèse, ce qui permet d'éviter les contraintes précédentes concernant les largeurs de bandes des filtres de synthèse. On détermine par la suite les caractéristiques du banc d'analyse par application de la loi (1) aux caractéristiques du banc de synthèse. Ceci permet d'éviter le problème du calcul de la racine carrée d'un nombre négatif.

La procédure de correction est maintenant organisée selon les étapes suivantes :

Étape 1 :

localisation de la fréquence du premier harmonique F_{a1} par un premier filtre passe bande du banc d'excitation. A sa sortie le signal alimente un limiteur délivrant un signal carré de fréquence F_{a1} et d'amplitude plus ou moins un selon le signe du signal d'entrée.

Étape 2 :

un filtre passe bande du banc d'analyse centré sur F_{h1} (obtenue en remplaçant dans (1) F_a par F_{a1}) permet après détection d'enveloppe d'estimer une amplitude moyenne du signal à la fréquence F_{h1} .

Étape 3 :

la modulation en amplitude du signal carré (étape 1) par l'amplitude moyenne détectée à l'étape 2 restituée à l'harmonique considéré sa véritable amplitude.

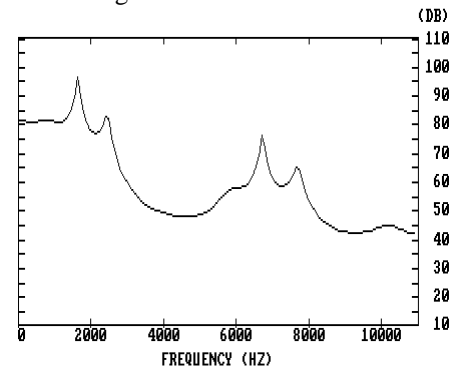
Étape 4 :

un filtre passe bande du banc de synthèse identique à celui de l'excitation de même indice élimine les harmoniques dus au limiteur.

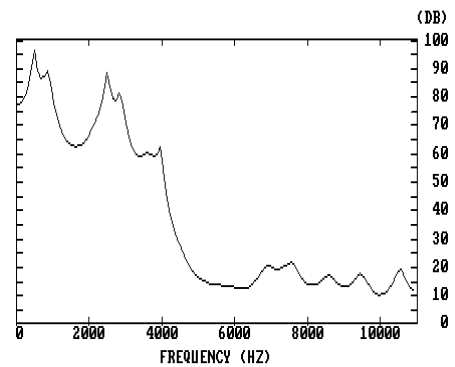
Étape 5 :

répétition des étapes 1, 2, 3 et 4 pour les harmoniques supérieurs.

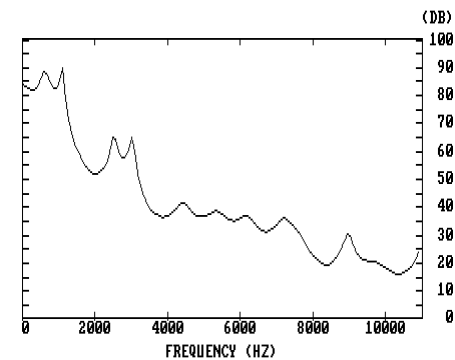
L'application de cette méthode à une voyelle /a/ prononcée par un locuteur masculin à une profondeur de 237 m avec un mélange d'hélium (hélium + oxygène) donne les résultats représentés sur la figure 2



(a) : enveloppe spectrale dilatée d'un son voisé /a/ "hyperbare"



(b) : enveloppe spectrale du son voisé /a/ "hyperbare corrigé"



(c) : enveloppe spectrale du son voisé /a/ "air" prononcé par le même locuteur

Figure 4 : Résultat du traitement de la voyelle /a/ prononcée par un locuteur masculin.

On peut observer la bonne restitution des fréquences des premiers maxima (donc des formants), ce qui est essentiel pour l'intelligibilité

2.3 Réduction du temps de traitement

Pour assurer un fonctionnement temps réel de notre correcteur, il est important de minimiser son temps de traitement. Ce dernier, dépend essentiellement du nombre de canaux N , de la vitesse de calcul dans les filtres numériques

et de la fréquence d'échantillonnage F_e . C'est en général, la fréquence d'échantillonnage qui fixe la largeur de la bande de traitement. La fréquence maximale F_{maxa} de la bande d'analyse B_{ana} doit être inférieure ou égale à la moitié de la fréquence d'échantillonnage ($F_{maxa} \leq F_e/2$ Shannon). Une fois B_{ana} fixée ($B_{ana} = F_{maxa} - F_{mina}$), le nombre de canaux N , n'est fonction que de l'échelonnement utilisé. La vitesse de calcul dans les filtres dépend quant à elle, du type de filtre utilisé, de la technique de synthèse et de l'algorithme de filtrage. C'est ce qui détermine le nombre M d'échantillons des différents filtres ou encore le nombre de multiplications à effectuer par seconde [1].

On peut observer en premier que dans notre vocodeur, la fréquence maximale du banc d'excitation F_{maxe} identique à la fréquence maximale du banc de synthèse F_{maxs} , est inférieure à $F_e/4$. Ceci, nous permet après élimination des composantes fréquentielles indésirables, supérieures à F_{maxs} par filtrage passe bas (filtre demi bande), de réduire de moitié la fréquence d'échantillonnage. Les calculs dans ces deux derniers bancs s'effectuent désormais à une cadence de $F_e/2$. L'ordre M des filtres d'excitation et de synthèse se trouve ainsi divisé par 2.

Les signaux à l'entrée des modulateurs doivent avoir la même fréquence d'échantillonnage, on doit donc réduire la fréquence d'échantillonnage par exemple à la sortie des différents filtres passe bas des canaux d'analyse, ce qui est possible puisque leur fréquence de coupure est très inférieure à $F_e/2$. Cette technique, permet un gain en temps de traitement de l'ordre de 43 % sans distorsion perceptible de la parole corrigée.

Pour réduire davantage le temps de traitement, on peut aussi alimenter les filtres passe bande des canaux d'analyse, dont les fréquences de coupure haute sont inférieures à $F_e/4$ (plus de $N/2$ canaux sont concernés), à partir du signal d'entrée de fréquence d'échantillonnage $F_e/2$. Pour les canaux d'analyse dont la fréquence de coupure est supérieure à $F_e/4$, un module formé d'un filtre passe bas suivi d'un réducteur de fréquence d'échantillonnage (décimateur) doit être intercalé entre le redresseur et le filtre passe bas de sortie des canaux concernés. Le gain obtenu en terme de réduction du temps de traitement est de l'ordre de 60 % et ceci sans aucune dégradation observable de la qualité de la parole corrigée.

3 Résultats

Les résultats obtenus avec cette nouvelle approche sont encourageants puisque pour les profondeurs où les comparaisons sont possibles (< 300m) un test d'écoute par sondage sur 23 personnes confirme l'amélioration due à cette technique en terme d'intelligibilité. Les phrases suivantes (issue d'un corpus français standard) :

- "Les deux camions se sont heurtés de face" (à - 80 m)
- "A l'ouest mes pommiers donnent peu" (à - 201m)
- "Un train entre déjà en gare" (à - 201m)
- " Le renard se hâte vers son gîte " (à - 80 m)

corrigées par vocodeur à canaux avec l'ancien et le nouvel algorithme ont été soumises d'une façon aléatoire aux 23 auditeurs. Les résultats du test d'écoute sont les suivants :

Préfèrent l'ancien vocodeur 19 (20.7 %)

Préfèrent le présent vocodeur 73 (79.3 %)

Pour ces tests, les écarts dépassent largement les 20,9 % d'incertitude statistique liée au nombre d'auditeurs. La nouvelle méthode de correction comparée à l'ancienne technique présente d'une part l'avantage de ne pas être limitée en profondeur et d'autre part de permettre le choix des largeurs de bandes des filtres de synthèse. Dans le précédant algorithme ces largeurs sont trop étroites surtout, aux grandes profondeurs, parce qu'elles sont imposées par celles du banc d'analyse. Ceci entraîne une distorsion de la parole corrigée et un effet de réverbération gênant pour l'ensemble des phrases traitées. Enfin, pour réduire le temps de traitement afin d'assurer un fonctionnement en temps réel du correcteur, on a adapté la fréquence d'échantillonnage à la bande utile qui change au cours du traitement, ce qui permet un gain en temps de traitement de l'ordre de 60 % sans distorsion perceptible de la parole corrigée.

4 Références

- [1] Bellanger M. "Traitement Numérique du Signal", 4ème édition révisée Masson 1990.
- [2] Calliope "La parole et son traitement automatique", Collection Technique et Scientifique des Télécommunications CNET/ENST, Masson 1989.
- [3] Fant G. and Lindqvist G. "Pressure and gaz mixture effects on diver's speech", Speech Transmission Laboratory, Royal Inst. Of Tech., Stockholm, Sweden, Quart. Prog. And Status Rep. STL-QPSR, pp. 7-17, 1968.
- [4] Fant G. and Lindqvist J. "Speech distorsion at high pressure in underwater physiology", Proc. 4th Int. Cong. Diver Physiology. New York : Academic, pp. 239-299, 1971.
- [5] Golden R.M. "Improving Naturalness and Intelligibility of Helium-Oxygen Speech, Using Vocoder Techniques", J.A.S.A., Vol 40, N°3, pp. 621-624, 1966.
- [6] Malherbe J.C. "About hyperbaric speech processing", Electronic Control & Measuring Systems 2nd workshop, Liberec, Czech Republic, pp. 114-122, 1995.
- [7] Morrow C.T. "Speech in deep submergence atmosphere", J.A.S.A. Vol. 50 n°3 part (1), pp. 715-728, 1971.
- [8] Saadane A. "Optimisation d'un vocodeur à canaux pour la correction de la parole hyperbare", Thèse Université de Rennes I, 1989.
- [9] Saadane A. et Malherbe J.C. "Optimisation des filtres d'un vocodeur à canaux pour la correction de la parole hyperbare", 1^{er} congrès Français d'Acoustique, Journal de physique, Colloque C2, suppl. N°2, Tome 51 pp. 793-796, 1990.
- [10] Saadane A. et Malherbe J.C. "Vocodeurs à canaux : nouvelle approche pour la correction de la parole hyperbare", Journal de physique, colloque C5, Vol 4, Mai 1994.
- [11] Smail Z. "Optimisation de la correction du signal de parole hyperbare par un vocodeur à canaux modifié", Thèse Université de Rennes I, mars 1997.
- [12] Tanaka R. and Nakatsui M. "Formant frequency shifts under high ambient pressures", Journal of the Radio Research Laboratories, Vol 21, N°105 pp. 261-267, 1974.