

# Vocodeurs à canaux : nouvelle approche pour la correction de la parole hyperbare

A. SAADANE et J.C. MALHERBE\*

LATI, I.R.E.S.T.E., La Chantrerie, CP. 3003, 44087 Nantes cedex 03, France

\* GRESIL, I.U.T. de Lannion, rue Edouard Branly, BP. 150, 22300 Lannion, France

**Résumé :** En plongée profonde, l'inhalation de mélanges synthétiques hyperbares résoud les problèmes physiologiques liés à l'air. Elle altère par contre le fonctionnement de la phonation. La parole produite dans ces conditions est inintelligible. L'approche décrite ici pour corriger une telle parole associe les avantages respectifs des vocodeurs à bande de base et à canaux, des difficultés apparaissant avec les seuls vocodeurs à canaux. La fréquence fondamentale, supposée inchangée, est élaborée au niveau de la synthèse à partir du signal hyperbare lui-même.

**Abstract :** Professional deep sea-diving implies that divers inhale synthetic hyperbaric mixtures to cope with physiological problems observed with air. This modifies the manner speech is produced. Speech under these conditions is altered and unintelligible. This paper puts forward an approach to make such speech understandable using the respective advantages of base-band vocoders and channel vocoders to deal with the difficulties arising from latter. The pitch (periodic glottic wave) supposed unaltered is obtained for synthesis from the hyperbaric signal itself.

## 1. INTRODUCTION

L'effet "Donald Duck" [1] caractérisant la parole hyperbare est dû, pour un mélange respiratoire synthétique donné, aux différences, par rapport à l'air atmosphérique, de sa masse volumique  $U$  et de sa célérité du son  $C$ . Les variations de composition relative du mélange avec la pression ambiante pour des raisons physiologiques rendent l'effet dépendant de la profondeur. Cette variation se traduit essentiellement par un déplacement des fréquences des formants dont une modélisation est donnée par la loi de Fant-Lindquist [2]

$$(F_h)^2 = (kF_a)^2 + (U_h/U_a - 1)(kF_o)^2$$

avec :  $F_a, F_h$  = fréquence d'un même formant respectivement dans l'air et dans le mélange hyperbare,  
 $k$  =  $C_h/C_a$  où  $C_h$  = vitesse du son dans le mélange hyperbare,  
 $C_a$  = vitesse du son dans l'air et à la pression atmosphérique,  
 $U_h$  = masse volumique du mélange hyperbare,  
 $U_a$  = masse volumique de l'air à la pression atmosphérique.  
 $F_o$  = fréquence de résonance du conduit vocal fermé aux lèvres (air, 1 atm). [3]

La composante non linéaire,  $(U_h/U_a - 1)(kF_o)^2$ , affectant les faibles fréquences peut être négligée en première approximation. Une simple compression d'un rapport  $k$  de l'enveloppe spectrale permet donc de restituer la position des formants et donc l'intelligibilité de la parole.

## 2. CHOIX DE LA STRUCTURE

Zurcher [4] réalisa, d'une manière analogique, cette compression par un vocodeur à canaux, formé d'un

analyseur et d'un synthétiseur.

L'analyseur est constitué d'une batterie de filtres passe-bande contigus couvrant la bande de fréquence transmise. Chaque filtre est associé à une détection filtrage passe-bas et à un quantificateur. Les fréquences de coupure des filtres passe-bas varient en général entre 20 et 50 Hz (au delà il ne semble pas y avoir un accroissement perceptible de la qualité). L'analyseur est formé en outre de la détection et de la mesure du pitch. L'information ainsi recueillie est aussi quantifiée et multiplexée avec celles fournies par les filtres.

La synthèse est effectuée par sommation des signaux délivrés par une batterie de filtres passe-bande contigus. Le nombre de filtres de synthèse est égal à celui d'analyse. La répartition des fréquences centrales et bandes passantes tient compte de la loi de compression envisagée. Chaque filtre reçoit un signal, périodique ou bruit blanc selon que le son est voisé ou non voisé, modulé en amplitude par la grandeur, proportionnelle à une énergie, issue du redressement filtrage qui suit le filtre d'analyse du même rang.

Une des difficultés à résoudre dans la conception de tels systèmes est la restitution de l'excitation (pitch et harmoniques). Plusieurs méthodes ont été proposées [5]. Une classification possible est donnée par Zurcher. Les algorithmes liés à toutes ces méthodes peuvent évidemment être appliqués au cas hyperbare si l'on suppose que le pitch de la parole hyperbare est inchangé.

On a adopté ici une structure de vocodeur à canaux modifié, ce que deux raisons au moins justifient :

- les critères de transmission bas débit différent des critères de correction de la parole déformée,
- l'élaboration de l'excitation à partir du fondamental transmis confère à la parole reconstruite un caractère subjectif "peu synthétique".

Cette structure s'inspire, dans la restitution du pitch, des vocodeurs à bande de base. Ces derniers transmettent intégralement la partie basse fréquence du signal de parole lui-même, signal qui va servir en réception à la reconstitution de l'excitation et de la bande basse.

Le synoptique général du correcteur numérique proposé est donné par la figure 1.

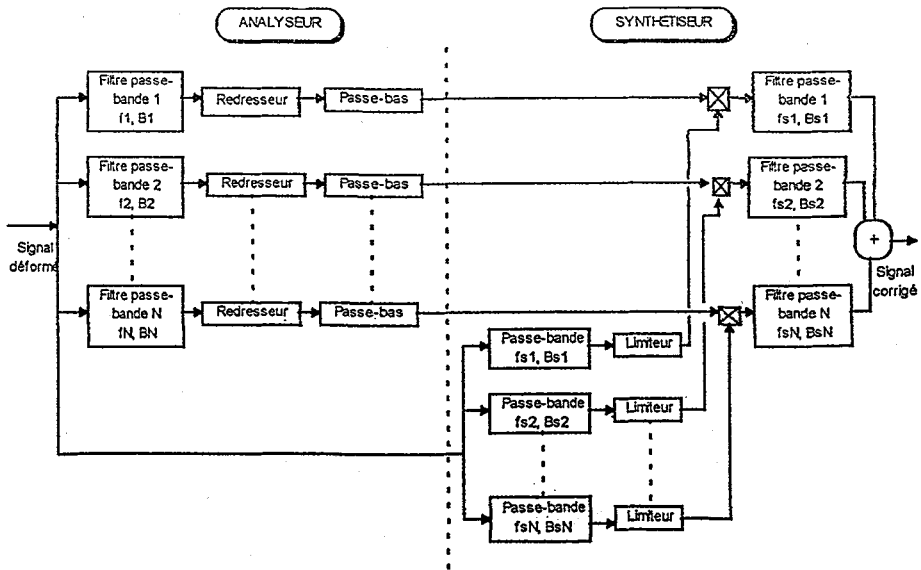


figure 1 - Schéma synoptique du vocodeur modifié

A l'analyse l'ensemble des canaux donne l'enveloppe spectrale du signal d'entrée dans la bande B. A la synthèse des filtres passe-bande contigus sont utilisés pour couvrir la bande B/k. La problématique détection du fondamental est ici remplacée par une troisième batterie de filtres, identique à celle de la synthèse. Ces filtres, suivis d'un limiteur dont la sortie est + 1 ou - 1 selon le signe de l'entrée, déterminent la structure fine de l'excitation dans la bande filtrée. La modulation en amplitude des harmoniques de cette excitation par le spectre d'analyse correspondant restitue le signal comprimé dans le rapport k.

### 3. OPTIMISATION DU VOCODEUR

L'optimisation de ce type de vocodeur passe par celle des filtres d'analyse et de synthèse. Une étude détaillée est donnée dans [6] et [7].

Rappelons que leur échelonnement se rapproche de celui des bandes critiques de l'oreille et que leur choix s'est fait sur la base de trois critères :

- les variations de l'intensité et de l'enveloppe spectrale des différents sons exigent une dynamique d'au moins 50 dB,
- la réponse fréquentielle composite des différents filtres doit avoir une amplitude constante et une phase linéaire,
- les caractéristiques de ces passe-bande doivent être un compromis entre une bonne résolution spectrale et une réponse transitoire rapide.

Le résultat obtenu, en exploitant à la fois la linéarité de la phase des filtres R.I.F. et la symétrie des régions de transitions résultant de l'approximation par une fenêtre, est une réponse fréquentielle présentant une ondulation maximale de 1,5 dB crête à crête. Cette ondulation est négligeable puisqu'elle se manifeste aux hautes fréquences là où le système auditif est peu sensible.

### 4. RESULTATS

Les données sont acquises et enregistrées sous forme de fichiers. Chaque fichier est constitué de "tranches" dont on peut définir le contexte (nombre de points).

La figure 2 présente 2 tranches de la voyelle |a| émise dans l'air à la pression atmosphérique.

La figure 3 présente 4 tranches de la même voyelle, prononcée par le même locuteur en atmosphère Héliox (He + O<sub>2</sub>) à une profondeur de 169 m. (pression ≈ 17,9 bars).

Une évaluation approximative de la périodicité temporelle des figures 2 et 3 nous permet de constater une variation du pitch (environ 150 Hz dans l'air et 165 Hz dans l'Héliox). Toutefois l'écart reste faible et le fait de considérer le pitch inchangé demeure une hypothèse acceptable.

La figure 4 représente les enveloppes spectrales lissées dans l'air et dans le milieu hyperbare. La comparaison de ces deux courbes autorise deux remarques :

- la non-linéarité du déplacement spectral est confirmée,
- en ne tenant pas compte de ce déplacement les deux enveloppes apparaissent similaires en forme à l'exception de la perte d'énergie (environ 8 dB) aux hautes fréquences.

Cette chute est principalement due aux facteurs physiologiques et s'expliquerait par une utilisation plus accentuée de la cavité nasale pour les grandes profondeurs. La figure 5 représente la correction linéaire apportée. La translation vers les basses fréquences est faite. Les valeurs trouvées pour les 2 premiers formants sont légèrement inférieures aux valeurs "air". Ceci se justifie parce que le déplacement de ces formants est

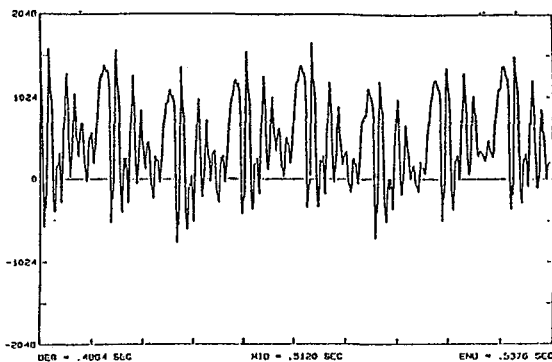


figure 2 : voyelle |a|, air

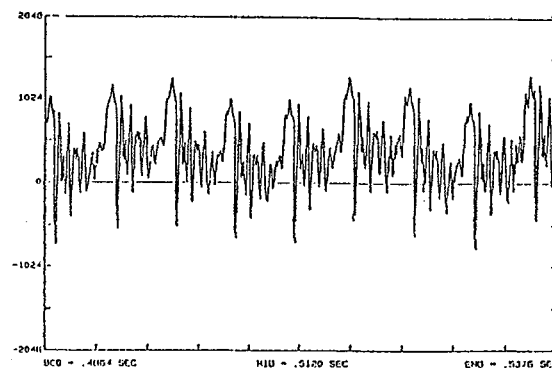


figure 3 : voyelle |a|, HélioX 169 m

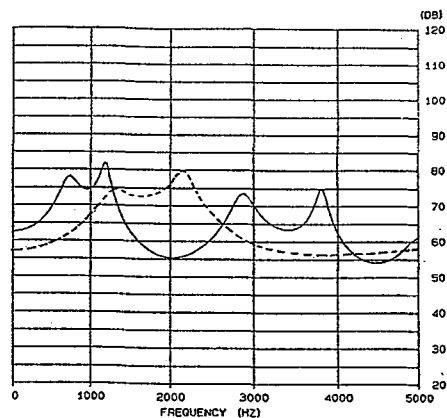


figure 4 : Enveloppe spectrale, voyelle |a|,  
— = air, - - - = hélioX 169 m

surtout affecté par l'effet de vibrations des parois du conduit vocal.

On note qu'en absence de tout dispositif corrigeant l'atténuation des hautes fréquences, celle-ci est évidemment conservée. Sachant que l'information essentielle des fricatives est contenue dans les hautes fréquences, on peut s'attendre a priori à une mauvaise correction de ces dernières. Une solution à ce problème consiste, soit :

- à placer à l'entrée du décodeur-correcteur un circuit analogique pour renforcer les fréquences élevées,
- à utiliser, en acquisition, un microphone ayant une préaccentuation  $\geq 12$  dB/oct en hautes fréquences.

**5. DISCUSSION ET CONCLUSION**

Outre la correction des mots le traitement d'une phrase permet de porter un jugement subjectif mais intéressant sur les transitions entre les différents sons. L'exemple choisi est : "les deux camions se sont heurtés de face". La parole restituée est, selon 6 auditeurs, "assez" intelligible et présente notamment peu l'effet synthétique observé avec d'autres types de correcteurs. Toutefois seuls des tests d'écoute sur plusieurs personnes permettraient une évaluation plus objective. Signalements enfin que sur la base des travaux et simulations effectués, la conception du décodeur est naturellement possible. La structure (actuellement à l'étude) est représentée en figure 6. Le principe consiste à insérer des microprocesseurs de traitement du signal (DSP) sur un bus VME. Le processeur maître est un 68000 de Motorola. Les processeurs esclaves sont des TMS 320 de Texas-Instruments. Chacun de ces esclaves sera implanté sur une carte au format "simple" ou "double-europe". Pour l'analyse-synthèse en temps réel trois cartes au moins sont nécessaires pour respectivement l'analyse, l'élaboration du pitch et la synthèse.

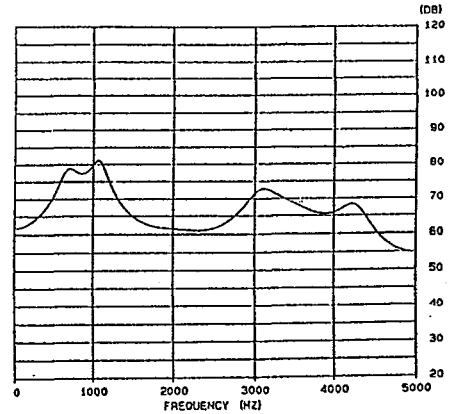


figure 5 : Enveloppe spectrale, voyelle | a |, corrigée

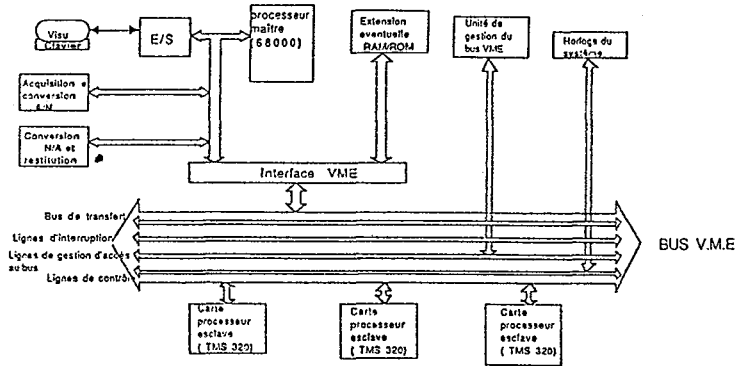


figure 6

**REFERENCES**

- [1] Richards M.A., "Helium speech enhancement using the short-time-fourier-transform", P.H.D. Georgia Institute of Technology, (1982).
- [2] Fant G., "Acoustic theory of speech production", Mouton, The Hague (1970).
- [3] Morrow C.T., "Speech in deep-submergence atmosphere", JASA Volume 50 n° 3 part (1) (1971).
- [4] Zurcher F., "Le transcoding "C.N.E.T." de la voix en atmosphère d'hélium", Note Technique TMA/ETA/24 janvier 74 (1974).
- [5] Hess W., "Pitch determination of speech signals", Springer-Verlag (1983).
- [6] Saadane A., "Optimisation d'un vocodeur à canaux pour la correction de la parole hyperbare", Thèse Université de Rennes 1 (France) (juillet 1989).
- [7] Saadane A. & Malherbe J.C., "Optimisation des filtres d'un vocodeur à canaux pour la correction de la parole hyperbare", Premier Congrès Français d'Acoustique, Colloque C2, supplément au n° 2, Tome51, (février 1990) pp. 793 - 796.