SUR LA QUASI-STATIONNARITE DU FILTRE VOCAL EN CONDITIONS "HYPERBARES"

J. CRESTEL, M. GUITTON, V. LE CALVE, M. CORAZZA

ENSSAT / LASTI
6, rue de Kérampont B.P. 447 22305 Lannion

RÉSUMÉ

D'une manière générale les techniques d'analyse, de codage, de traitement du signal de parole sont fondées sur des modélisations du système vocal encadrées par un ensemble d'hypothèses, dont la quasi-stationnarité du filtre vocal et la quasi-périodicité de l'excitation glottique. Cette communication démontre, à partir d'une caractérisation théorique de la réponse impulsionnelle du conduit vocal, la validité a fortiori de ces hypothèses dans le cas de la parole "hyperbare". Les termes de l'application, à ce type de signal, de la transformée de Fourier à court terme et de la modélisation AR sont précisés en conséquence.

ABSTRACT

Generally speaking, methods for analysis, coding and processing of speech signal are based on vocal system models, along with a set of hypothesises. The quasi-stationnarity of the vocal filter and the quasi-periodicity of the glottal excitation are the two main points in this set. This paper proves above all the validity of these hypothesises for hyperbaric speech signal from a theoretical characterization of the vocal tract impulse response. Consequently the terms for applying the TFCT to such signals are detailed.

I-INTRODUCTION

Les plongeurs professionnels qui réalisent des plongées en saturation sont contraints, physiologiquement, d'inhaler des mélanges respiratoires synthétiques (héliox, hydrox, hydréliox) où les gaz hélium et hydrogène constituent les diluants majoritaires de l'oxygène. Conséquemment la parole produite - dite "hyperbare"- est inintelligible. Il est prouvé que ce phénomène résulte de l'étalement spectral du signal de parole induit par la modification de la fonction de transfert du conduit vocal [1]. En outre l'analyse expérimentale révèle les caractères suivants, comparativement au signal de parole "air" :

-une augmentation de la fréquence fondamentale [2]

-une diminution de la vitesse d'élocution [3]

-une atténuation de l'énergie des sons non voisés relativement à celle des sons voisés.

L'amélioration de l'intelligibilité est l'objet de recherches depuis une vingtaine d'années. Les travaux récents ont révélé deux méthodes qui s'avèrent être les plus performantes, et néanmoins perfectibles moyennant une optimisation de leur mise en oeuvre. L'une utilise la transformée de Fourier à court terme (TFCT) [4], l'autre est fondée sur les propriétés de la prédiction linéaire [5]. Les deux méthodes consistent en un traitement numérique par blocs de durée constante θ. Par principe l'information globale contenue dans le bloc est considérée significative de la représentation temps-fréquence du signal quel que soit l'instant t dans l'intervalle θ. Cette théorie a fait l'objet, dans le cadre de l'application de la TFCT, d'une formulation fondée sur le modèle linéaire de production de la parole [6], (fig.1).

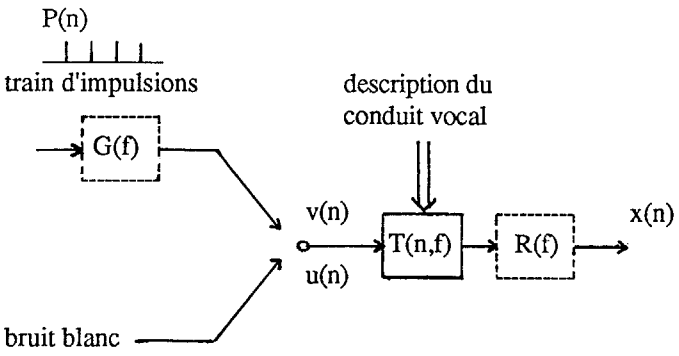

fig.1 Modèle linéaire de production de la parole.

La fonction de transfert T(n,f) intègre les filtres "glotte" G(f) et "rayonnement aux lèvres" R(f). T(n,f) est la transformée de Fourier de la réponse du système vocal à une impulsion appliquée à l'instant n-m, soit

$$T(n,f) = \sum_{m=-\infty}^{+\infty} t(n,m)\, e^{-j2\pi fm}$$

v(n) est un train d'impulsions quasi-périodique dont la forme harmonique s'écrit :

$$v(n) = \frac{1}{P(n)} \sum_{k=0}^{P(n)-1} e^{jk[\varnothing(n)+\varnothing_0]}$$

et où $\phi(n)$ désigne la phase instantanée du signal.
Le bruit blanc stationnaire centré u(n) est caractérisé par son second moment σ_u et sa fonction d'autocorrélation

$$R_u(\tau) = \sigma_u^2 \, \delta_0(\tau)$$

Conformément à ce modèle, la TFCT est une représentation temps-fréquence du signal de parole réaliste sous réserve que pendant la durée de la réponse impulsionnelle t(n,m) les variations du pitch P(n) d'une part et des paramètres définissant la configuration articulatoire d'autre part soient négligeables (hypothèses de quasi-périodicité de l'excitation impulsionnelle et de quasi-stationnarité du filtre vocal [6]). Trois cas sont explicités dans [4].

-P(n) et T(n,f) constants : la TFCT du signal donne des harmoniques de même forme et d'amplitude proportionnelles aux échantillons de T(n,f).

-P(n) constant, variation linéaire de T(n,f) : toutes les harmoniques ont même forme mais leurs amplitudes sont biaisées.

-T(n,f) constant, variation linéaire de P(n) : les harmoniques ont des formes différentes.

En pratique, selon [4,6], et s'agissant du signal de parole "air" les deux hypothèses (quasi-périodicité et quasi-stationnarité) sont réalistes si θ n'excède pas 30 ms. S'agissant du signal de parole "hyperbare" qu'en est-il de la validité de ces hypothèses? La réponse que nous apportons dans cette communication repose sur une étude de la réponse impulsionnelle du conduit vocal et prend en compte les caractères de la parole "hyperbare" énumérés précédemment.

II-REPONSE IMPULSIONNELLE DU CONDUIT VOCAL EN CONDITIONS HYPERBARES

D'une manière générale la caractérisation de la réponse impulsionnelle du conduit vocal peut être expérimentale (mesures directes ou indirectes), théorique (étude de comportement d'un modèle du conduit vocal), mixte (analyse du signal de parole sur la base d'un modèle de fonctionnement du conduit vocal) : les résultats présentés relèvent de l'approche théorique.

L'étude est fondée sur une modélisation n-tubes du conduit vocal [7] régie par les équations de propagation du son de Webster et intègre des paramètres relatifs à :

-la configuration articulatoire (fonction d'aire).

-la caractérisation du mélange respiratoire : masse volumique ρ, célérité du son c, coefficient de viscosité μ, constante adiabatique η, coefficient de conduction de chaleur λ, chaleur spécifique à pression constante Cp.

-la prise en compte des phénomènes de pertes justifiant les bandes passantes des formants : transfert de chaleur, friction, vibration des parois, rayonnement aux lèvres.

La fonction de transfert échantillonnée T(m) exprime le rapport des débits aux niveaux des lèvres et de la glotte. Elle est définie en fonction des paramètres pour des domaines de variation homogènes [0..fmax] de la variable fréquence, l'homogénéité étant dictée par la loi de Fant :

$$f_h^2 = \left(\frac{c_h}{c_a} f_a\right)^2 + \left(\frac{\rho_h}{\rho_a} - 1\right) \left(\frac{c_h}{c_a} f_{wa}\right)^2$$

avec

-f_h, f_a : fréquences de formants correspondants "hyperbare" et "air".

-f_{wa} : fréquence de résonance du conduit vocal fermé aux lèvres.

-c_h, c_a : célérités du son en milieu "hyperbare" et en milieu "air".

- ρ_h, ρ_a : masses volumiques du mélange respiratoire en milieu "hyperbare" et en milieu "air".

Pratiquement l'application de la loi est généralisée à tout f de l'intervalle de définition "air" et permet en particulier d'associer à la borne f_a max de l'intervalle "air" une borne f_h max d'un intervalle "hyperbare".

La réponse impulsionnelle t(k) est calculée par une transformation de Fourier inverse de T(m). Les résultats présentés correspondent à une configuration de conduit vocal uniforme (voyelle [ə]). La comparaison d'une réponse impulsionnelle "air" et d'une réponse impulsionnelle "hydrox 250m" (fig.2) met en évidence une différence sensible tant de la structure harmonique que de l'amortissement.

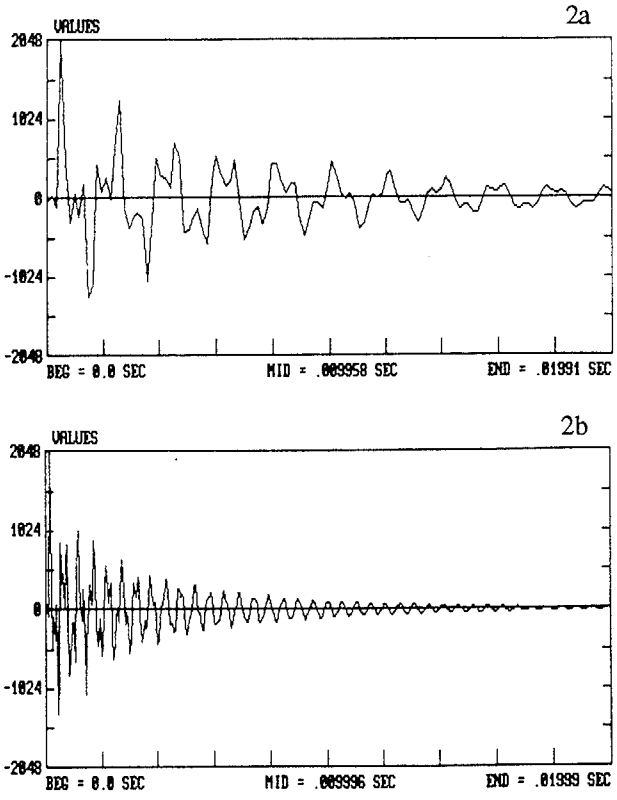

fig.2 Réponses impulsionnelles "air" (2a), "hydrox" (2b).

Sachant que seul le facteur d'amortissement est significatif pour l'étude et afin d'introduire une quantification du

phénomène, il est judicieux d'analyser la réponse en termes d'énergie normalisée E(k) telle que

$$E(k) = \frac{\sum_{i=k}^{\infty} t^2(i)}{\sum_{i=0}^{\infty} t^2(i)}$$

et de définir sa longueur l à partir de la relation E(l) = 0.05.

Selon cette approche l est une fonction des 6 variables ρ, c, μ, η, λ, Cp. En fait les mélanges respiratoires hyperbares synthétiques sont caractérisés par une pression partielle d'oxygène constante. Les proportions des gaz constituants sont alors fonction de la pression totale, et sachant que les paramètres physiques du mélange peuvent être déterminés avec une bonne précision à partir de ceux des constituants, la grandeur l peut être vue comme fonction de la seule variable profondeur Pr. La figure 3 traduit la sensibilité de l à Pr dans le cas des quatre mélanges respiratoires air, héliox, hydrox, hydréliox.

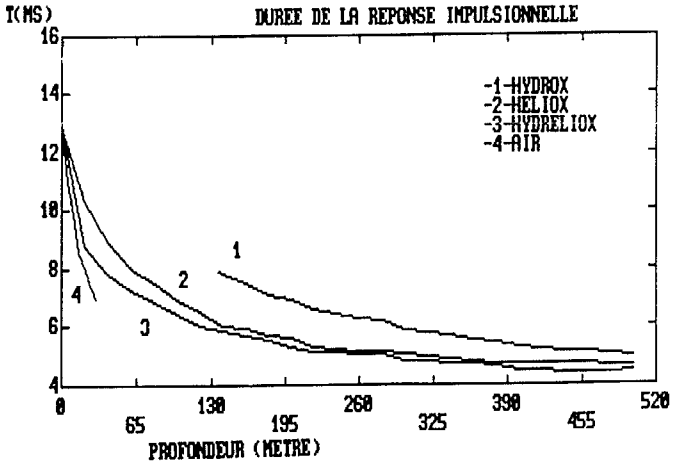

fig.3 Longueur de la réponse impulsionnelle en fonction de la profondeur pour les mélanges respiratoires air,héliox, hydrox, hydréliox.

L'allure des courbes incite, plutôt qu'à faire un distinguo, à interpréter globalement la réduction de l au-delà de 150m quel que soit le mélange respiratoire. Ce phénomène confère nécessairement des propriétés spécifiques au signal de parole "hyperbare". On notera que l'air n'est pas tolérable physiologiquement au-delà de 60m, que l'hydrox n'est pas utilisable techniquement en-deça de 150m.

III-ANALYSE COMPARATIVE DES PROPRIETES DES SIGNAUX DE PAROLE "AIR" ET "HYPERBARE"

La quasi-périodicité traduit l'invariance relative du pitch pendant la durée de la réponse impulsionnelle. La grandeur l.ΔP(n) en est donc significative, ΔP(n) symbolisant une variation locale du pich. En supposant ΔP(n) indépendant du mélange respiratoire, la réduction de l.ΔP(n) induite par l (de l'ordre de 3) renforce l'hypothèse de quasi-périodicité de l'excitation voisée.

La quasi-stationnarité traduit l'invariance relative de la fonction de transfert T(n,f) pendant la durée de la réponse impulsionnelle. En corrélant la variation temporelle de la configuration articulatoire, autrement dit ΔT(n,f), à la vitesse d'élocution Vel (telle que ΔT(n,f) / Vel=constante), l'expression l.ΔT(n,f) / Vel peut être considérée représentative de la quasi-stationnarité. La quasi-stationnarité est d'autant plus réaliste que la valeur de l'expression est plus faible. Dans le cas de l'inhalation d'un mélange respiratoire hyperbare la réduction sensible à la fois de Vel [3] (dans un rapport de l'ordre de 1.2) et de l (de l'ordre de 3) conforte nettement l'hypothèse de quasi-stationnarité du filtre vocal.

IV-IMPLICATIONS PRATIQUES

4-1-Transformation de Fourier du signal "hyperbare"

Le caractère de quasi-stationnarité autorise l'analyse du signal de parole voisé par TFCT. L'analyse en bande étroite exige une fenêtre d'observation de durée θ suffisante. Concrètement la bande passante B du filtre d'analyse (fenêtre) doit vérifier B<1/P(n). Pour une fenêtre de hamming B vérifie B=4/θ. Deux contraintes antagonistes sont donc imposées : θ>4/P(n) et quasi-stationnarité. Or la période de pitch "hyperbare" est plus faible que le pitch "air" [2], ce qui autorise à finesse d'analyse égale, une fenêtre de durée θ plus faible. Cet aspect favorable, et les meilleures quasi-stationnarité et quasi-périodicité intrinsèques associées au signal de parole "hyperbare" font de celui-ci un meilleur candidat que le signal de parole "air" à l'analyse et au traitement par l'outil TFCT.

4-2-Modélisation AR

Considérons le modèle linéaire de production de la parole et soit y(n) le signal défini par

$$Y(z) = X(z) / R(z) \,.\, G(z)$$

ou de manière équivalente, en introduisant la réponse impulsionnelle t(n,m) du conduit vocal

$$y(n) = \sum_{m=-\infty}^{+\infty} u(m) \,.\, t(n,n-m)$$

Assimilons l'excitation impulsionnelle à un peigne de Dirac,

$$u(m) = \sum_{k=-\infty}^{+\infty} \delta(m - kP)$$

il vient

$$y(n) = \sum_{k=-\infty}^{+\infty} t(n,\, n - kP) \qquad (1)$$

Soit V(n,z) le filtre "conduit vocal" tel que

$$V(n,z) = T(n,z) / R(z) \,.\, G(z)$$

En modélisant le conduit vocal par un filtre tout-pôle et en admettant l'hypothèse de quasi-stationnarité (justifiée dans le §III) il vient

$$V(n,z) = 1 / A(z)$$

$$\text{avec} \quad A(z) = \sum_{i=0}^{M} a_i \, z^{-i}$$

Les coefficients a_i caractérisent une réponse impulsionnelle h(m)

$$h(m) = - \sum_{i=1}^{M} a_i \, h(m-i)$$

avec $h(0) = 1$, $h(m-i) = 0$ pour $m-i < 0$
L'identification des coefficients a_i, autrement dit de h(m), est fondée sur la connaissance partielle de la fonction d'autocorrélation $R_y(k)$ de y(n). Le critère classique de minimisation de l'erreur quadratique totale optimalise, à un facteur près, l'estimation $R_h(k) = \hat{R}_y(k)$ pour $0 <= k <= M$. Or compte-tenu de la relation (1) et de l'inégalité $l < P$ propre au contexte hyperbare, y(n) est obtenu par périodisation de la réponse impulsionnelle t(n,m), sans enchevêtrement:
donc

$$R_t(k) = R_y(k) \text{ pour } k < P$$

On en déduit que h(m) et T(n,z) sont, dans le contexte hyperbare, des représentations pertinentes du comportement intrinsèque du conduit vocal, l'ordre M étant supposé judicieusement choisi. En outre la modélisation AR s'avère être un support privilégié pour l'analyse expérimentale du signal, complémentaire à l'analyse théorique (§II).

V-CONCLUSIONS

La majorité des méthodes d'analyse et de transformation du signal hyperbare fondent leur justification sur une simple extension des propriétés de quasi-stationnarité et de quasi-périodicité caractérisant le signal "air". Il est montré que, s'agissant du signal de parole "hyperbare", ces propriétés sont plus réalistes et se traduisent par des utilisations théoriquement mieux fondées de la transformée de Fourier et de la prédiction linéaire. Ce résultat est d'un intérêt immédiat pour l'élaboration de méthodes de restitution de l'intelligibilité de la parole "hyperbare" par traitement numérique du signal. En particulier la prédiction linéaire constitue un outil privilégié compte-tenu de la pertinence de la discrimination source-conduit vocal par filtrage inverse.

BIBLIOGRAPHIE

[1] J. Crestel, M. Guitton
Application de la modélisation de l'appareil phonatoire à la caractérisation de la parole hyperbare
Congrès Français d'Acoustique Lyon, avril 1990

[2] H.Hollien, W.Shearer, J.W.Hicks
Voice fundamental frequency levels of divers in helium-oxygen speaking.
Undersea Biomedical Research, Vol. 4, n° 2, pp 199-207, june 1977

[3] Nakatsui M., Suzuki J., Takasugi T., Tanaka R.
Nature of helium speech : systematic observation and analysis during simulated dives.
2nd Int. Ocean Dev. Conf., Tokyo, pp 1615-1627, 1972

[4] M.A. Richards
Helium speech enhancement using the short-time Fourier transform.
Ph.D Georgia Institute of Technology, 1982

[5] J. Crestel, M. Guitton
Un système pour l'amélioration des communications en plongée profonde
Colloque Gretsi, Nice, juin 1987

[6] M.R. Portnoff
Short-time Fourier analysis of sampled speech
IEEE Transactions on acoustics, speech, and signal processing
Vol. ASSP-29, N°3, pp 364-373, june 1981

[7] J.L Flanagan
Speech analysis, synthesis, and perception, second edition.
New-York : Springer-Verlag, 1976