

Codage robuste des paramètres LSF du codeur de parole FS1016 : Transmissions à travers un canal bruité

Merouane BOUZID et Bachir BOUDRAA

Laboratoire Communication Parlée et Traitement du Signal (LCPTS)
Faculté d'Electronique et d'Informatique, Université *USTHB*, BP 32, El-Alia,
Bab-Ezzouar, ALGER, 16111, ALGERIE. Email: mbouzid@yahoo.com

ABSTRACT

In this paper, we present a channel optimized trellis coded vector quantization system designed for effective and robust coding of LSF spectral parameters. The aim of this encoding system (designed at the beginning for noiseless channel) is to achieve a low bit rate transparent quantization of LSF parameters of the US Federal Standard (FS1016) speech coder. Once the effectiveness of our FS1016 LSF's encoder was proven in the case of ideal transmissions, we were interested after in the improvement of its robustness for transmissions over noisy channel. To implicitly protect the transmission indices of the LSF's encoder incorporated in the FS1016, we used a joint source-channel coding carried out by the channel optimized vector quantization method.

1. INTRODUCTION

Dans les systèmes de codage de la parole à débit réduit, l'information spectrale à court terme d'un signal de parole est souvent modélisée par la réponse fréquentielle d'un filtre tout-pôles $H(z) = 1/A(z)$ avec $A(z) = 1 + a_1 z^{-1} + \dots + a_{10} z^{-10}$. Les dix coefficients de prédiction linéaire (LPC) $\{a_i\}_{i=1,\dots,10}$ sont pratiquement inappropriés à la quantification. Parmi les représentations équivalentes qui se sont avérées plus efficaces, les fréquences de raies spectrales LSF (Line Spectral Frequencies) [1] sont sans doute les plus utilisées. Exploitant les propriétés connues des paramètres spectraux LSF [2], divers schémas de codage ont été développés dans le passé pour la quantification efficace de ces paramètres.

Dans cet article, nous présentons un système de codage robuste des paramètres LSF, conçu à base d'un quantificateur vectoriel codé par treillis optimisé (OTCVQ). Le but de ce système, appelé au début "Encodeur LSF-OTCVQ", est de réaliser une quantification transparente à bas débit des paramètres LSF. Dans le cas des transmissions à travers un canal non bruité, on a déjà prouvé dans [3] que l'encodeur LSF-OTCVQ (avec distance pondérée) pouvait réaliser de bonnes performances lors du codage des LSF du codeur de parole FS1016 [4]. Par la suite, on s'est intéressé à l'amélioration de la robustesse de l'encodeur LSF-OTCVQ pour des transmissions à travers un canal bruité. Conventionnellement, un codage de canal redondant [5] est utilisé pour assurer une protection "explicite" aux paramètres sensibles des codeurs de parole contre les erreurs de canal. Selon l'approche de la conception

séparée, le codeur de canal peut être conçu d'une façon complètement séparée du codeur source en ajoutant des bits de redondance aux paramètres à protéger [5]. Cette approche séparatrice permet bien de concevoir des systèmes de codage robustes mais au prix de l'augmentation des débits/délais de transmission et de la complexité du codage/décodage. Cependant, à bas débit où les contraintes en complexité et en délais de traitement sont très sévères, ce genre de codage de canal "séparé" redondant n'est pas forcément recommandé. Pour assurer une protection implicite (c.-à-d., sans redondance) des données source tout en maintenant un débit constant et une complexité réduite, les chercheurs ont introduit le dit codage conjoint source-canal (CCSC) dans le quel on se préoccupe de la minimisation de la distorsion totale en considérant simultanément l'impact des erreurs de canal et de la distorsion du codage de source [6], [7].

Pour protéger implicitement notre encodeur LSF-OTCVQ incorporé dans le FS1016, nous avons adopté une méthode de CCSC relative au quantificateur vectoriel (VQ) conçu en tenant compte de l'erreur de canal. Il s'agit de la quantification vectorielle optimisée pour un canal bruité (COVQ) où les opérations de codages de source et de canal sont intégrées conjointement dans une seule entité COVQ en incorporant le paramètre du canal dans le processus de conception [7].

2. QUANTIFICATION VECTORIELLE OPTIMISEE POUR UN CANAL BRUITE

Un quantificateur vectoriel optimisé pour un canal bruité (COVQ) est un schéma de codage qui se base sur le principe de généralisation du VQ conventionnel en tenant compte du bruit du canal. Rappelons qu'un VQ, de dimension k et de débit R bits/échantillon (bpe), est défini comme une fonction d'un espace euclidien \mathfrak{R}^k vers un dictionnaire fini $Y = \{y_0, \dots, y_{L-1}\}$ composé de $L = 2^{kR}$ vecteurs-code y_i [8]. La conception du dictionnaire VQ consiste à décomposer l'espace des vecteurs source x en L classes disjointes $\{R_0, \dots, R_{L-1}\}$ (partition) et à associer à chaque classe R_i un représentant (centroïde) y_i telle que la distorsion moyenne totale D soit minimisée [8]. La plus populaire des méthodes utilisées pour concevoir un VQ optimal est sans doute l'algorithme LBG [9]. Cet algorithme est une application itérative des deux conditions nécessaires d'optimalité (c.-à-d., les règles du plus proche voisin et du centroïde) telle que la partition et le dictionnaire soient mis à jour itérativement [8]. Pour un canal bruité, la conception du système de

codage COVQ est réalisée par une version étendue au cas bruité de l'algorithme LBG de la quantification vectorielle (LBG-VQ) [7], [6]. L'extension réside seulement dans la formulation des conditions nécessaires d'optimalité et de la distorsion moyenne totale en incluant la probabilité d'erreur du canal. En effet, l'optimisation COVQ est réalisée en utilisant une expression généralisée de la distorsion moyenne totale, formulée par [6], [7] :

$$D = \frac{1}{k} \sum_{i=0}^{L-1} \int_{R_i} p(x) \left[\sum_{j=0}^{L-1} p(j/i) \cdot d(x, y_j) \right] dx, \quad (1)$$

où $p(j/i)$ est la probabilité conditionnelle du canal, $p(x)$ est la distribution de la source et $d(x, y_j) = \|x - y_j\|^2$ est la mesure de distance quadratique moyenne (MSE) entre un vecteur source x et son représentant y_j . Dans nos conceptions, nous avons considéré que le canal de transmission est un canal discret de type binaire symétrique (BSC) de probabilité de transition p [5], [10].

Les étapes de conception du COVQ restent les mêmes que celles du VQ; il s'agit toujours d'une optimisation itérative des deux conditions d'optimalité "généralisées" en minimisant la distorsion moyenne totale généralisée. Pour $i = 0, \dots, L-1$, ces conditions d'optimalité sont formulées ci-dessous. Les classes R_i sont mises à jour suivant la règle généralisée du plus proche voisin :

$$R_i = \left\{ x \in R^k : \sum_{j=0}^{L-1} p(j/i) \|x - y_j\|^2 \leq \sum_{j=0}^{L-1} p(j/l) \|x - y_j\|^2, \forall l \neq i \right\} \quad (2)$$

Les vecteurs-code y_j du dictionnaire optimisé COVQ sont actualisés suivant la condition généralisée du centroïde :

$$y_j = \frac{\sum_{i=0}^{L-1} p(j/i) \int_{R_i} xp(x).dx}{\sum_{i=0}^{L-1} p(j/i) \int_{R_i} p(x).dx}, \quad j = 0, \dots, L-1 \quad (3)$$

Les étapes de notre version de l'algorithme COVQ sont détaillées dans [10]. La plus grande difficulté dans la conception d'un tel système est que la probabilité d'erreur du canal BSC est un paramètre d'entrée dans la procédure COVQ. L'estimation de ce paramètre (noté ϵ et appelé probabilité d'erreur de conception) doit être proche de la probabilité réelle p du canal BSC lors des transmissions.

Des exemples de résultats de simulation de systèmes COVQ, conçus pour différentes valeurs du paramètre de conception ϵ , sont tabulés ci-dessous. Ces systèmes COVQ (caractéristiques: $k = 2$ et $R = 2$ bpe) ont été appliqués au codage de sources gaussiennes sans mémoire pour des transmissions à travers un canal bruité de type BSC de probabilité d'erreur p variable.

Table 1 : Evaluation comparative des performances-RSB des systèmes COVQ et VQ opérant à travers canal BSC

$\epsilon \backslash p$	0.000	0.001	0.005	0.01	0.05
0.000	9.686	9.685	9.624	9.537	8.664
0.001	9.584	9.604	9.565	9.481	8.643
0.005	9.292	9.314	9.357	9.332	8.571
0.01	8.927	8.965	9.034	9.179	8.477
0.05	6.824	6.918	7.351	7.608	7.800
0.1	4.650	5.292	5.875	6.801	7.043

Dans le cas des transmissions à travers un canal de probabilité p élevée, ces résultats montrent que les performances des systèmes COVQ sont nettement supérieures à celles du VQ standard conçu par LBG-VQ pour un canal non bruité (c.-à-d., $\epsilon = 0.000$). Cependant, lorsque le canal est non bruité ($p = 0.000$) ou faiblement bruité les performances-RSB des systèmes COVQ se dégradent avec l'augmentation du paramètre ϵ .

3. CODAGE ROBUSTE DES LSF DU FS1016 : APPLICATION DE L'ENCODEUR LSF-OTCVQ

Pour réaliser une quantification transparente des paramètres LSF du codeur FS1016, nous avons conçu un système de codage basé sur une version optimisée de la quantification vectorielle codée par treillis (OTCVQ) [3].

La OTCVQ, rappelons le brièvement, est une technique de quantification qui améliore les systèmes de codage en treillis traditionnel en étiquetant les branches du treillis avec des sous-ensembles d'un dictionnaire-TCVQ étendu et optimisé [11]. Pour quantifier une séquence de vecteurs-source d'entrée, l'algorithme de Viterbi est utilisé pour trouver la meilleure séquence possible des vecteurs-code (chemin optimal) à travers le treillis qui minimise la distorsion totale. A la fin du codage, l'encodeur OTCVQ transmet à la réception : une séquence de bits indiquant le chemin optimal, une séquence d'indices binaires des vecteurs-code choisis à l'intérieur des sous-ensembles du chemin optimal et des bits spécifiant l'état initiale du treillis. Plus de détails sur le principe de conception des encodeurs TCVQ/OTCVQ et leurs performances peuvent être trouvés dans [11], [3], [10].

En utilisant donc la technique OTCVQ, un schéma d'encodage, appelé au début "Encodeur LSF-OTCVQ", a été conçu pour obtenir une quantification transparente des paramètres LSF à des débits plus bas. En général, une quantification de qualité transparente est obtenue si les trois conditions suivantes sont maintenues [2]: 1)- la distorsion spectrale (SD) moyenne est d'environ 1 dB, 2)- le pourcentage des trames externes (outliers frames) ayant une SD entre 2 et 4 dB est moins de 2% et 3)- aucune trame "outliers" ne doit avoir une SD qui dépasse 4 dB. Pour coder un vecteur LSF de taille 10, le treillis de l'encodeur LSF-OTCVQ est composé de 5 étages où chaque étage est utilisé pour quantifier une paire 2-D du

vecteur LSF. Ainsi, 5 dictionnaires-OTCVQ 2-D étendus sont conçus pour coder un vecteur LSF.

Dans le cas des transmissions idéales à travers un canal non bruité, nous avons montré dans [3] que nos encodeurs LSF-OTCVQ, conçus avec une mesure de distance pondérée, pouvaient réaliser des quantifications-LSF de qualité transparente à des débits plus bas comparés à ceux conçus avec une mesure de distance conventionnelle non pondérée. Notons que la mesure de distance pondérée, que nous avons utilisée, été formulée par [2], [3] :

$$d(f, \hat{f}) = \sum_{i=1}^{10} c_i w_i (f_i - \hat{f}_i)^2, \quad (4)$$

où f_i et \hat{f}_i sont respectivement les $i^{\text{ème}}$ coefficients des vecteurs LSF original f et quantifié \hat{f} ; c_i et w_i représentent respectivement les poids fixe et variable assignés au $i^{\text{ème}}$ coefficient du vecteur LSF.

D'autre part, les résultats des évaluations objective et subjective présentés aussi dans [3] et [10] ont montré que l'encodeur LSF-OTCVQ de 27 bits/trame, incorporé dans le FS1016, pouvait assurer une quantification de qualité transparente de ses paramètres LSF avec en plus un gain de 7 bits/trame. En effet, les paramètres LSF du standard fédéral FS1016 de 4.8 Kbits/s [4] sont codés à l'origine par un SQ non uniforme de 34 bits/trame. Aussi, en terme de RSB segmental moyen, une amélioration significative des performances-RSBseg objectives du FS1016 a été obtenue par application de l'encodeur LSF-OTCVQ.

3.1 Robustesse de l'encodeur LSF-OTCVQ pour des transmissions par canal bruité

Pour protéger implicitement les indices de transmission de l'encodeur LSF-OTCVQ de 27 bits/trame, incorporé dans le FS1016, nous avons adopté une méthode de CCSC par COVQ. Le principe de conception du système LSF-OTCVQ optimisé pour un canal bruité, que nous avons dénoté par "Encodeur COVQ-LSF-OTCVQ" [10], se base principalement sur l'algorithme de conception de l'encodeur LSF-OTCVQ modifié et réadapté selon le concept de base de la COVQ.

Les principales modifications ont été apportées au niveau du calcul des distances et des vecteurs-code optimisés pour un canal bruité. En effet, les équations (3) et (5) ont été combinées pour définir une mesure de distance appropriée entre un vecteur LSF original et sa version quantifiée. Cette distance a été formulée comme suit [10]:

$$d(f, \hat{f}_i) = \sum_{j \in \xi_i} p(j/i) \frac{1}{k} \sum_{m=1}^k c_m w_m d(f(m) - \hat{f}_j(m))^2, \quad (5)$$

où $k = 2$ pour les paires du vecteur LSF et ξ_i est l'ensemble des voisins de l'indice i qui contient tous les indices j telle que la distance de Hamming entre i et j soit égale à un. Précisons que dans nos applications, les 5

dictionnaires étendus de l'encodeur COVQ-LSF-OTCVQ ont été optimisés pour une probabilité de conception $\epsilon = 0.05$. La base d'apprentissage utilisée dans la conception était composée de 75000 vecteurs LSF construits à partir d'environ 43 minutes de parole prise de la base de données internationale TIMIT [12]. Le nombre d'états du treillis de l'encodeur est $S = 4$; par conséquent 2 bits par trame sont nécessaires pour spécifier l'état initial du treillis lors du codage. Les 25 bits restants sont repartis comme suit : 5 bits/trame pour le chemin optimal et 20 bits/trame pour les indices des vecteurs-code des 5 dictionnaires étendus selon l'allocation (5 5 4 4 2).

Nous présentons maintenant des exemples de résultats de simulation portant sur les performances du système d'encodage COVQ-LSF-OTCVQ ($\epsilon = 0.05$) de 27 bits/trame appliqué au codage efficace et robuste des paramètres LSF du FS1016. Pour évaluer la robustesse de cet encodeur, les indices de ses vecteurs-code sont transmis volontairement à travers un canal bruité BSC de probabilité d'erreur p variant entre 0 et 0.5. En utilisant une base de données d'environ 13.69 secondes de parole extraite de la base de test TIMIT, des versions synthétiques de signaux de parole ont été générées par le codeur FS1016, avec des évaluations des performances des encodeurs LSF incorporés simultanément et affectés par les mêmes séquences de bruit. Les évaluations comparatives des performances-SD des encodeurs de 27 bits/trame: LSF-OTCVQ (sans protection) et le COVQ-LSF-OTCVQ ($\epsilon = 0.05$) sont reportées dans la table 2.

Ces résultats montrent que lorsque la probabilité d'erreur de transmission devient assez élevée ($p > \epsilon = 0.05$), la COVQ apporte une amélioration significative aux performances de l'encodeur LSF-OTCVQ de base. Sans protection, le LSF-OTCVQ a subi une dégradation plus sévère comparée au système protégé. Ceci est traduit par une augmentation brutale de la SD moyenne du LSF-OTCVQ quand le canal devient fortement bruité, ainsi que le pourcentage des fenêtres "outliers" ayant une SD > 4 dB qui commence à s'accroître à des probabilités d'erreur p faible. Dans ces conditions de transmission, l'encodeur COVQ-LSF-OTCVQ a permis d'avoir une bonne robustesse contre les erreurs de canal en maintenant une augmentation réduite et lente de la SD moyenne et du nombre des "outliers" de SD > 4 dB. Par ailleurs, lorsque les transmissions s'effectuent à travers un canal faiblement bruité ($p \leq \epsilon$), les performances du COVQ-LSF-OTCVQ deviennent sous-optimales; compromettant ainsi les performances de quantification transparente (en terme de SD et de fenêtres "outliers"). Cependant, des constatations très importantes ont été notées concernant les performances objectives en terme de RSBseg moyen du codeur de parole FS1016. En effet, à l'inverse de certaines remarques notées à la section 2, l'amélioration des performances RSBseg du FS1016 (avec LSF codés par COVQ-LSF-OTCVQ) est aussi remarquable lorsque le canal de transmission est faiblement bruité (table 3). Le système global présente ainsi une bonne robustesse contre les erreurs du canal bruité.

Table 2 : Comparaison des performances entre les encodeurs COVQ-LSF-OTCVQ et LSF-OTCVQ

Probabilité p du BSC	Encodeur COVQ-LSF-OTCVQ 27 bits/trame ($\epsilon = 0.05$)			Encodeur LSF-OTCVQ 27 bits/trame (sans protection)		
	SD Moy. (dB)	SD "Outliers" (en %)		SD Moy. (dB)	SD "Outliers" (en %)	
		2-4 dB	> 4 dB		2-4 dB	> 4 dB
0.000	1.690	25.607	0.441	1.073	0.440	0.000
0.001	1.693	25.827	0.441	1.099	0.441	0.442
0.005	1.710	26.710	0.662	1.148	0.883	1.544
0.050	1.800	32.671	0.883	1.707	10.596	7.505
0.100	1.924	38.852	0.883	2.696	15.010	21.192
0.200	2.130	46.799	3.532	4.251	17.439	43.929
0.500	2.696	67.911	7.726	6.649	13.245	80.573

Table 3 : Performances objectives du FS1016 global

Probabilité p du BSC	RSBseg moyen [dB] du FS1016 avec :	
	LSF-OTCVQ (27 bits/trame)	COVQ-LSF-OTCVQ (27 bits/trame)
0.000	10.499	10.497
0.001	10.491	10.490
0.005	10.472	10.481
0.050	10.146	10.324
0.100	9.721	10.246
0.200	8.668	10.140
0.500	6.958	9.847

5. CONCLUSION

Dans ce travail, un système d'encodage optimal basé sur la technique OTCVQ a été développé pour le codage efficace et robuste des paramètres LSF du codeur de parole FS1016. L'objectif était d'améliorer la robustesse de l'encodeur LSF-OTCVQ conçu au début pour encoder les paramètres LSF du FS1016 opérant à travers un canal idéal non bruité. Les résultats de simulation ont montré que notre nouveau système d'encodage COVQ-LSF-OTCVQ ($\epsilon = 0.05$) de 27 bits/trame a apporté une bonne protection implicite lors du codage des paramètres LSF du codeur FS1016 opérant à travers un canal bruité (BSC). En effet, il a permis au système LSF-OTCVQ de base d'avoir une bonne robustesse contre les erreurs de canal surtout lorsque la probabilité des erreurs de transmission est assez élevée. Par contre, lorsque les transmissions s'effectuent à travers un canal non bruité ou faiblement bruité, les performances-SD du système COVQ-LSF-OTCVQ deviennent sous-optimales. Cependant, les performances objectives RSBseg de l'encodeur global FS1016 (avec LSF codés par COVQ-LSF-OTCVQ) restent aussi satisfaisantes dans ces conditions de transmission.

BIBLIOGRAPHIE

- [1] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", Journ. of Acoust. Society America, vol.57, p.535, Apr. 1975.
- [2] K. K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", IEEE Trans. on Speech and Audio Processing, Vol. 1, no. 1, pp. 3-14, Jan. 1993.
- [3] M. Bouzid, A. Djeradi et B. Boudraa, "Optimized Trellis Coded Vector Quantization of LSF Parameters: Application to the 4.8 Kbps FS1016 Speech Coder", Signal Processing, Vol. 85, Issue 9, pp. 1675-1694, Sept. 2005.
- [4] J. P. Campbell, T. E. Tremain and V. C. Welch, "The Proposed Federal Standard 1016 4800 bps Voice Coder: CELP", Speech Technology Magazine, pp. 58-64, April/ May 1990.
- [5] L. Shu, *An Introduction to Error-Correcting Codes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA, 1970.
- [6] S. B. Z. Azami, P. Duhamel, O.Rioul, Combined Source-Channel Coding: Panorama of Methods, Presented in CNES Workshop on Data Compression, Toulouse France, Nov. 1996.
- [7] N. Farvardin and V. Vaishampayan, "On the performance and Complexity of Channel-Optimized Vector Quantizers", IEEE, Trans. on Infor. Theory, Vol.37 n°.1, pp. 155-159, Jan. 1991.
- [8] A. Gersho, R. M. Gray, *Vector quantization and Signal compression*, Kluwer Academic Publishers, USA 1992.
- [9] Y. Linde, A. Buzo and R. M. Gray, "An Algorithm for Vector Quantization Design", IEEE Trans. on Communications, Vol. COM-28, pp.84-95, Jan. 1980.
- [10] M. Bouzid, "Codage conjoint de source et de canal pour des transmissions par canaux bruités", Thèse de Doctorat, spécialité Communication parlée, université USTHB, Alger, 2006.
- [11] T.R. Fischer, M.W. Marcellin & M. Wang, "Trellis-coded vector quantization", IEEE Trans. on Infor. Theory, Vol.37, pp.1551-1566, Nov. 1991.
- [12] J. S. Garofolo and al., "DARPA TIMIT Acoustic-phonetic Continuous Speech Database", Technology Building, National Institute of Standards and Technology (NIST), Gaithersburg, October 1988.