

Codage efficace à bas débit des paramètres spectraux LSF

M. BOUZID, A. DJERADI, B. BOUDRAA

Laboratoire Communication Parlée et Traitement du signal
Equipe Codage de la Parole, Faculté d'Electronique et d'Informatique
Université USTHB, BP 32, El-Alia, Bab-Ezzouar, ALGER, 16111, ALGERIE
Fax de la faculté: 213-21- 24.76.07. Email: mbouzid@yahoo.com

ABSTRACT

Speech coders operating at low bit rates necessitate efficient encoding of the linear predictive coding (LPC) coefficients. Line spectral Frequencies (LSF) parameters are currently one of the most efficient choices of transmission parameters for the LPC coefficients. In this paper, an optimized trellis coded vector quantization (TCVQ) scheme for encoding the LSF parameters is developed. When the selection of a proper distortion measure is the most important issue in the design and operation of the encoder, an appropriate weighted distance measure has been used during the TCVQ construction process. Using this distance, we will show that our LSF TCVQ encoder performs better than the encoder conceived with the unweighted distance.

1. INTRODUCTION

Dans des applications de codage de la parole, l'information spectrale à court terme d'un signal de parole est souvent modélisée par la réponse fréquentielle d'un filtre tous-pôles de fonction de transfert $H(z) = 1/A(z)$, avec $A(z) = 1 + a_1z^{-1} + \dots + a_{10}z^{-10}$. Les 10 coefficients $\{a_i\}_{i=1,2,\dots,10}$ de ce filtre autorégressif, connus sous le nom de coefficients de prédiction linéaire (LPC) [1], sont dérivés du signal d'entrée en utilisant une analyse par prédiction linéaire (LP) sur chaque trame du signal parole. Dans la pratique, on ne quantifie pas directement les coefficients LPC car ils ne sont pas appropriés au codage. Plusieurs transformations équivalentes ont été développées afin de les convertir en paramètres beaucoup plus appropriés à la quantification. Parmi les représentations qui se sont avérées efficaces, les fréquences de raies spectrales LSF (Line Spectral Frequencies) sont sans doute les plus utilisées [2]. Les paramètres LSF, qui sont liés aux zéros du polynôme $A(z)$, présentent un certain nombre de propriétés intéressantes [1, 3]. Exploitant ces propriétés, divers schémas de codage basés sur la quantification scalaire et vectorielle ont été suggérés pour la quantification efficace des paramètres LSF. Les schémas de codage à base d'un quantificateur scalaire (SQ) sont intéressants dûs à leur niveau bas de complexité; cependant, ils accomplissent la qualité de quantification transparente à des débits hauts [4]. Un quantificateur vectoriel (VQ) peut réaliser la qualité de quantification transparente à des débits binaires plus bas [3]. Cependant, il est plus complexe et exige des

tailles mémoire élevées. Pour réduire la complexité des calculs et les exigences en taille mémoire, divers schémas à base de VQs sous-optimaux, comme les VQs multi-étages [5], les VQs divisés (Split) [3],..., ont été proposés dans le passé pour coder les paramètres LSF.

Dans cet article, nous présentons un système de codage optimisé à quantification vectorielle codée par treillis TCVQ (Trellis Coded Vector Quantization) pour encoder les paramètres LSF à bas débit. Dans ce système d'encodage, que nous avons appelé LSF-OTCVQ, la dépendance intra-trame entre les plus proches paires de paramètres LSF (deux voisins) sera exploitée en utilisant des dictionnaires bidimensionnels (2-D). Connaissant que le choix d'une mesure de distorsion appropriée est une question importante dans la conception et le fonctionnement d'un système VQ, nous avons étudié la meilleure mesure de distance qui doit être utilisée pendant l'étape de conception de notre encodeur LSF-OTCVQ.

2. QUANTIFICATION VECTORIELLE CODEE PAR TREILLIS

La quantification scalaire codée par treillis (TCQ) est une forme améliorée du codage en treillis traditionnel qui attribue aux branches du treillis des sous-ensembles (ou sous dictionnaires) de niveaux de quantification plutôt que des niveaux de quantification individuels [6,7,8]. L'approche TCQ, qui a été motivée par la formulation d'Ungerboeck de la Modulation Codée par Treillis (TCM) [9], utilise un dictionnaire (alphabet) structuré avec un ensemble augmenté de niveaux de quantification.

Pour coder une source sans mémoire de dimension $k=1$ par un codeur TCQ opérant à un débit de R bits/échantillon (bpe), on peut utiliser n'importe quel treillis d'Ungerboeck pour modulation d'amplitude [9]. L'alphabet augmenté (doublé) peut être choisi comme l'ensemble des niveaux de quantification obtenus par un SQ de Lloyd-Max de débit $R+1$ bpe. Ainsi, un alphabet se composant deux fois d'autant de grandeurs scalaires, à savoir 2^{kR+1} ($k=1$), est construit. Cependant, lors du processus de codage, la structure du treillis ramène le nombre augmenté de niveaux de quantification au débit de codage désiré (R). Par conséquent, seulement 2^{kR} de ces niveaux peuvent être utilisés pour représenter un échantillon de source à un instant donnée. Basé sur les règles de la partition d'ensemble d'Ungerboeck [9], ces

niveaux sont ensuite divisés en 4 sous-ensembles (D_0, D_1, D_2, D_3) et étiquetés aux branches du treillis. Pour quantifier la séquence source, le célèbre algorithme de Viterbi est utilisé pour trouver le chemin optimal à travers le treillis. La séquence des niveaux choisis par l'algorithme de Viterbi est représentée par une séquence de bits indiquant le chemin optimal (séquence de sous-ensembles des niveaux de quantification) en plus d'une séquence de bits (mots-code) nécessaire pour indiquer les niveaux choisis à l'intérieur des sous-ensembles du chemin optimal. A la réception, le décodeur reconstruit la source quantifiée comme suit : la séquence de bits indiquant le chemin sélectionné à travers le treillis est utilisée comme entrée du codeur convolutionnel du système TCQ; la sortie de ce codeur sélectionne les sous-ensembles appropriés. Les mots-code de la seconde séquence sont utilisés pour indiquer les niveaux corrects à l'intérieur de chaque sous-ensemble choisi. Le système de codage TCQ est complètement défini par la structure du treillis, les niveaux de quantification et la partition d'ensemble. Un exemple de codage d'une source uniformément distribuée sur l'intervalle $[-A, A]$ par un codeur TCQ à 4 états opérant à un débit $R=2$ bpe est illustré à la figure 1.

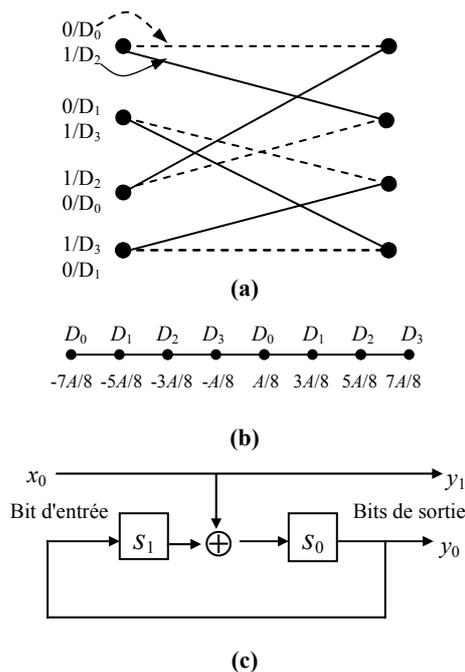


Figure 1 : Encodeur TCQ de débit $R=2$ bits/échantillon, (a) Section du treillis étiqueté d'Ungerboeck à 4 états, (b) Niveaux doublés de l'alphabet TCQ et la partition, (c) Codeur convolutionnel TCQ correspondant.

Bien que les performances d'un codeur TCQ [6, 10] sont dans la plupart des cas proches de la limite théorique $D(R)$, des améliorations sont toujours possible en généralisant sa structure au cas vectoriel. On parle alors de la quantification vectorielle codée par treillis (TCVQ). Pour un même débit de codage, la TCVQ fournit des résultats meilleurs que la TCQ [7, 8, 10]. En outre, elle

permet un codage à des débits fractionnaires, chose que la TCQ n'autorise pas. La structure d'un codeur TCVQ est similaire à celle d'un système TCQ, avec une légère augmentation de la complexité des calculs due à la recherche de vecteurs dans les sous-ensembles. Une meilleure conception du dictionnaire-TCVQ initial est réalisée par l'algorithme LBG d'un VQ (LBG-VQ) [11]. Une fois que les vecteurs-code sont déterminés, ils sont divisés dans des sous-ensembles et les sous-ensembles sont étiquetés aux branches du treillis suivant les mêmes règles de la partition d'ensemble d'Ungerboeck [9]. La partition d'ensemble qui est plutôt simple et rapide dans le cas scalaire, n'est pas une tâche facile dans le cas vectoriel. Dans la conception de nos encodeurs TCVQ, nous avons utilisé l'heuristique décrite dans [8] pour réaliser la partition d'ensemble du dictionnaire étendu.

Des exemples de résultats de simulation, obtenus lors du codage TCVQ de sources sans mémoire Gaussiennes, sont listés dans table 1. Pour différents débits fractionnaires, les résultats sont donnés en termes de rapport signal sur bruit (RSB) en dB, avec les performances du VQ-LBG conventionnel et les fonctions Distorsion-Débit $D(R)$ correspondantes.

Table 1 : Performances RSB de la TCVQ (à des débits fractionnaires) d'une source Gaussienne.

Débit (bpe)	Dim. k	Nombre d'états							VQ LBG	$D(R)$
		4	8	16	32	64	128	256		
0.66	6	3.34	3.39	3.41	3.42	3.45	3.48	3.49	3.03	4.01
0.75	4	3.72	3.78	3.80	3.82	3.87	3.90	3.93	3.35	4.51
0.80	5	3.96	4.04	4.07	4.08	4.14	4.18	4.20	3.70	4.82

Pour un même débit de codage, ces résultats de simulation montrent que les performances de la TCVQ sont nettement supérieures à celles du VQ-LBG standard. Comparé aux limites théoriques $D(R)$, d'éventuelles améliorations sont toujours possibles.

2.1. Optimisation de la TCVQ

Pour améliorer d'avantage les performances de la TCVQ, une procédure d'optimisation de la conception du dictionnaire-TCVQ a été élaborée. Cette procédure d'optimisation par entraînement utilise l'étape standard de l'algorithme LBG [11] appliquée à une séquence d'apprentissage de la source à coder. Elle consiste à mettre à jour le dictionnaire TCVQ en remplaçant chaque vecteur-code avec la moyenne de tous les vecteurs source attribués à ce vecteur. Ceci mène à un algorithme itératif de conception pour l'encodeur TCVQ global. En fait, chaque itération est un processus de codage TCVQ avec de nouveaux vecteurs-code optimisés, obtenus de l'itération précédente. En utilisant cette variante d'optimisation, l'algorithme sera dénoté par algorithme OTCVQ (Optimized Trellis Coded Vector Quantization). Des exemples de résultats de simulation obtenus lors du codage de sources sans mémoire gaussiennes, en utilisant nos encodeurs TCVQ avec dictionnaires optimisés, sont tabulés ci-dessous.

Table 2 : Performances de la TCVQ optimisée (à des débits fractionnaires) d'une source Gaussienne.

Débit (bpe)	Dim. k	Nombre d'états							VQ LBG	$D(R)$
		4	8	16	32	64	128	256		
0.66	6	3.41	3.45	3.47	3.48	3.49	3.52	3.53	3.03	4.01
0.75	4	3.81	3.85	3.87	3.89	3.93	3.96	3.97	3.35	4.51
0.80	5	4.08	4.14	4.16	4.17	4.21	4.23	4.25	3.70	4.82

En comparant ces résultats avec ceux donnés dans la table 1, nous remarquons bien l'amélioration des performances apportée par l'optimisation des dictionnaires de la TCVQ.

3. CODAGE DES LSF PAR TCVQ OPTIMISEE

En utilisant la technique de codage OTCVQ décrite précédemment, le schéma d'encodage LSF-OTCVQ et ses performances sont présentés dans cette section. Pour des applications de codage de la parole, la TCVQ est utilisée en mode bloc, où chaque bloc correspond à un vecteur LSF de taille 10. Dans notre travail, des dictionnaires 2-D sont utilisés pour coder les vecteurs LSF. Ainsi, chaque étage dans le treillis est associé à deux dimensions du vecteur LSF. Par conséquent, le treillis du système d'encodage LSF-OTCVQ est composé de 5 étages. Puisque les paramètres LSF ont des moyennes et des variances disparates, différents dictionnaires doivent être utilisés pour chaque étage. Ainsi, 5 dictionnaires 2-D étendus correspondant aux 5 étages du treillis sont nécessaires pour coder un vecteur LSF. Ces dictionnaires sont conçus d'abord par l'algorithme LBG-VQ [11]. Une fois la structure du treillis, les sous-ensembles et l'étiquetage des branches ont été conçus, les dictionnaires sont ensuite optimisés pour la base de données (vecteurs LSF) d'apprentissage, en utilisant l'algorithme OTCVQ décrit dans la section précédente.

Les performances du quantificateur sont évaluées par la distorsion spectrale moyenne SD (Spectral Distorsion) qui est souvent utilisée comme mesure objective de la performance d'encodage des paramètres LSF. Calculée discrètement sur une largeur de bande limitée, l'expression de la SD pour une trame i est donnée en décibels par [5]:

$$SD_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{n=n_0}^{n_1-1} \left[10 \log_{10} \frac{S(e^{j2\pi n/N})}{\hat{S}(e^{j2\pi n/N})} \right]^2} \quad (1)$$

Pour un signal de parole échantillonné à 8 kHz avec une largeur de bande de 3 kHz, une FFT de $N=256$ points est utilisée pour calculer les spectres de puissance, originaux et quantifiés, $S(e^{j2\pi n/N})$ et $\hat{S}(e^{j2\pi n/N})$ du filtre de synthèse LPC de la $i^{\text{ème}}$ trame du signal de parole. Ainsi, la SD est calculée avec une résolution de 31.25 Hz par échantillon. Dans [3], Paliwal et Atal ont établi que la SD moyenne n'est pas suffisante pour mesurer seule la qualité perçue. Selon leurs résultats, un codage- LPC transparent est obtenu si les trois conditions suivantes sont maintenues: 1)- la SD moyenne est d'environ 1 dB, 2)- le pourcentage

des trames "outliers frames" ayant une SD entre 2 et 4 dB est moins de 2% et 3)- aucune trame ne doit avoir une SD qui dépasse 4 dB.

La base de données parole utilisée dans les simulations se compose d'environ 43 minutes de parole prise de la base de données TIMIT [12]. Les signaux de parole sont d'abord filtrés passe-bas à une fréquence 3.2 KHz, puis sous échantillonnés à 8 KHz. Pour construire la base de données des vecteurs LSF, une analyse LP, d'ordre 10 par la méthode d'autocorrélation, est effectuée sur chaque fenêtre d'analyse de 30 ms (pondérée par la fenêtre de Hamming). Une partie de la base de données (75000 vecteurs LSF) est utilisée pour l'apprentissage et la partie restante, de 11262 vecteurs LSF (différente de la base d'apprentissage), est utilisée pour les tests.

Nous évaluons, à présent, les performances de notre système OTCVQ utilisé pour la quantification des LSF. Deux mesures de distorsion différentes ont été testées séparément dans la conception et le fonctionnement du LSF-OTCVQ. Les résultats de simulation reportés ici ont été obtenus en utilisant des treillis d'Ungerboeck à 4 états. Pour différents débits de codage, les performances de l'encodeur optimisé LSF-OTCVQ, en terme de SD moyenne et de fenêtres "outliers", sont montrées dans la table 3. Ces résultats ont été obtenus en utilisant une mesure de distance euclidienne non pondérée dans l'étape de conception et de fonctionnement de l'encodeur.

Table 3 : Performances de l'encodeur LSF-OTCVQ en fonction du débit de codage.

Bits/trame	SD Moy. (dB)	SD Outliers (en %)	
		2-4 dB	> 4 dB
24	1.34	7.04	0.03
25	1.24	3.97	0.03
26	1.18	3.01	0.02
27	1.14	2.95	0.02
28	1.04	1.60	0.01

Ces résultats montrent que l'encodeur LSF-OTCVQ, avec la mesure de distance Euclidienne non pondérée, a besoin d'environ 28 bits/trame (ou de plus) afin de réaliser un codage de quantification transparente.

3.1 Mesure de distance pondérée

Afin d'obtenir une quantification transparente à des débits plus bas, une autre mesure de distorsion plus appropriée a été choisie. Il s'agit de la mesure de distance euclidienne pondérée. Basé sur les principales propriétés des paramètres LSF, certaines formules de distances pondérées ont été proposées pour le codage des LSF [3, 13]. Si f et \hat{f} sont respectivement les deux vecteurs original et quantifié des paramètres LSF, alors la distance euclidienne carrée pondérée entre ces deux vecteurs est donnée par [3] :

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

où c_i et w_i représentent respectivement les poids fixe et variable assignés au $i^{\text{ème}}$ coefficient LSF. Plusieurs fonctions de pondération ont été définies pour calculer le vecteur des poids variables $w = [w_1, \dots, w_{10}]$. Dans notre conception, nous avons utilisé la fonction de pondération proposée dans [13]. Elle définie par :

$$w_i = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i}, \quad (3)$$

avec $w_0 = 0$ et $w_{11} = 0.5$.

Le vecteur additionnel de poids constants $c = [c_1, \dots, c_{10}]$ a été introduit dans la formule afin que les LSF des basses fréquences soit pondérés plus que les autres. Ce vecteur est déterminé expérimentalement [3] :

$$c_i = \begin{cases} 1.0, & \text{for } 1 \leq i \leq 8 \\ 0.8, & \text{for } i = 9 \\ 0.4, & \text{for } i = 10 \end{cases} \quad (4)$$

Pour différents débits de codage, les performances du système d'encodage LSF-OTCVQ, utilisant une mesure de distance euclidienne pondérée sont tabulées ci-dessous.

Table 4 : Performances de l'encodeur LSF-OTCVQ en utilisant une mesure de distance pondérée

Bits/ trame	SD Moy. (dB)	SD Outliers (en %)	
		2-4 dB	> 4 dB
24	1.29	5.26	0.02
25	1.19	2.99	0.00
26	1.15	2.72	0.00
27	1.07	1.90	0.00
28	0.98	1.10	0.00

En comparant ces résultats à ceux donnés dans la table 3, nous pouvons constater que la mesure de distance pondérée améliore les performances de l'encodeur LSF-OTCVQ en termes de SD moyenne et de nombre de fenêtres "outliers". Nous avons besoin ici de 27 bits/trame pour obtenir une qualité de quantification transparente.

4. CONCLUSION

Dans ce travail, un système basé sur la quantification vectorielle codée par treillis a été appliqué avec succès dans le codage efficace des paramètres LSF. En utilisant une mesure de distance pondérée dans la conception et l'opération de notre encodeur LSF-OTCVQ, une quantification de qualité transparente peut être réalisée à des débits plus bas. Comparé à l'encodeur conçu avec une distance non pondérée, l'LSF-OTCVQ avec la distance

pondérée peut diminuer le débit d'environ 1-2 bits/trame, tout en maintenant des performances comparables.

BIBLIOGRAPHIE

- [1] Kleijn (W.B.), Paliwal (K.K.), Speech coding and synthesis, *Elsevier Science B.V.*, 1995.
- [2] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", *Journ. of Acoust. Society of America*, vol.57, p.535, 1975.
- [3] K. K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3-14, Jan. 1993.
- [4] F. K. Soong and B. H. Juang, "Optimal quantization of LSP parameters," *Proc. IEEE Int. Conf. Acous., Speech Signal Processing*, New York, pp. 394-397, April 1988.
- [5] W. F. LeBlanc, B. Bhattacharya, S. A. Mahmoud & V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding", *IEEE Trans. Speech and Audio Processing*, Vol. 1, No. 4, pp. 373-385, October 1993.
- [6] M.W. Marcellin & T.R. Fischer, "Trellis coded quantization of memoryless and Gauss-markov sources ", *IEEE Trans. on Communication*, Vol. 38, pp. 83-93, January 1990.
- [7] T.R. Fischer, M.W. Marcellin & M. Wang, "Trellis-coded vector quantization", *IEEE Transactions on Information Theory*, Vol. 37, pp. 1551-1566, Nov 1991.
- [8] H.S. Wang & N. Moayeri, "Trellis coded vector quantization", *IEEE Trans. on Communication*, Vol. 40, pp. 1273-1276, August 1992.
- [9] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part I and II", *IEEE Commun. Magazine*, Vol.25, pp.5-21, Feb. 1987.
- [10] M. BOUZID, "Trellis Coded Vector quantization", Technical research report, speech coding team, Centre de Développement des Technologies avancées (CDTA), Alger, 2001.
- [11] Y.Linde, A.Buzo & R.M.Gray, "An Algorithm for Vector Quantization Design", *IEEE Trans. on Comm.*, Vol. COM-28, pp.84-95, Jan. 1980.
- [12] J. S. Garofolo et al., "DARPA TIMIT Acoustic-phonetic Continuous Speech Database", NIST, Gaithersburg, October 1988.
- [13] R. Laroia, N. Phamdo & N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers", *Proc. IEEE Int. Conference Acoustic Speech and Signal Processing*, pp. 641-644, May 1991.