

Codage efficace à débit variable basé sur la quantification vectorielle à divisions commutées: Application aux paramètres ISF en large bande

SALAH-EDDINE CHERAIATIA, MEROUANE BOUZID

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

cher.salah@yahoo.fr, mbouzid@usthb.dz

Résumé - Le codage efficace des coefficients de prédiction linéaire (LPC) est l'un des problèmes importants dans la conception des codeurs de parole modernes. Les paramètres "fréquences spectrales d'immitance" ISF (Immittance Spectral Frequencies) sont actuellement classés parmi les choix les plus appropriés pour représenter les coefficients LPC en large bande. Dans cet article, nous proposons une version à débit variable du quantificateur vectoriel à divisions commutées (SSVQ) développé pour le codage efficace des paramètres ISF de parole en large-bande, selon des suppositions de transmission à travers un canal non bruité.

Abstract - Modern speech coders necessitate efficient coding of the linear predictive coding (LPC) coefficients. Immittance Spectral Frequencies (ISF) parameters are currently the most efficient choices of transmission parameters for the LPC coefficients in wideband. In this paper, we propose a variable rate version of the switched split vector quantizer (SSVQ) scheme developed for efficient coding of wideband speech ISF parameters under noiseless channel conditions.

1 Introduction

Dans la plupart des systèmes de codage de la parole modernes, l'enveloppe spectrale à court terme d'un signal de parole est souvent modélisée par la réponse fréquentielle d'un filtre tout-pôle dont la fonction de transfert est donnée par $H(z) = 1/A(z)$, avec $A(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p}$ [1, 2]. Les coefficients de ce filtre $\{a_i\}_{i=1,2,\dots,p}$, connus aussi sous le nom de coefficients de prédiction linéaire (LPC), sont dérivés du signal d'entrée par une analyse par prédiction linéaire d'ordre p sur chaque trame du signal de parole. La quantification efficace des coefficients LPC joue un rôle important dans la préservation de l'intelligibilité et de la qualité naturelle du signal de parole codé sur toute la largeur de bande. En pratique, ces coefficients ont des propriétés de quantification médiocres. Ainsi, plusieurs représentations paramétriques équivalentes ont été formulées afin de les convertir en paramètres beaucoup plus appropriés à la quantification. Les fréquences de raie spectrales LSF (Line Spectral Frequencies) et les fréquences spectrales d'immitance ISF (Immittance Spectral Frequencies) se sont avérées les plus efficaces pour représenter les coefficients LPC des codeurs de parole modernes basés sur le modèle autorégressif.

Dans ce travail, on s'est intéressé particulièrement à la quantification efficace des paramètres ISF des codeurs de parole en large-bande comme le codeur large-bande à débit multiple adaptatif AMR-WB (Adaptive Multi-Rate Wide-Band) norme ITU-T G.722.2 [3]. Notons que les paramètres LSF et ISF sont des représentations équivalentes qui se déduisent mathématiquement les unes des autres par de simples fonctions de conversion. Les paramètres ISF (ISFs), qui

ont été introduits par Bistriz et al. [4], sont définies comme étant les pôles et les zéros d'une fonction d'immitance à la glotte :

$$I_{16}(z) = \frac{A(z) - z^{-16} A(z^{-1})}{A(z) + z^{-16} A(z^{-1})} \quad (1)$$

On obtient alors 16 paramètres ISF et le seizième est un coefficient de réflexion.

Comparé aux codeurs de parole en bande étroite (300–3400 Hz), les codeurs en large-bande ont amélioré le naturel et l'intelligibilité de la parole décodée en élargissant la bande passante utilisée pour la transmission du signal de parole (50–7000 Hz). Cependant, ils requièrent un nombre plus élevé de paramètres LPC, typiquement 16, pour représenter l'enveloppe spectrale de parole. Ainsi, le quantificateur vectoriel (VQ) conventionnel doit fonctionner à des débits plus élevés et sur des vecteurs de dimensions plus grandes si on veut l'utiliser pour le codage des ISFs. Ces exigences pratiques vont conduire à une augmentation excessive de la complexité des calculs et de la taille mémoire.

Dans le passé, divers schémas de quantificateurs vectoriels (VQs) structurés ont été développés pour le codage des paramètres ISF en large-bande. Ces schémas peuvent réduire considérablement la complexité des calculs avec une perte modérée des performances de quantification. L'un des schémas de codage les plus présentés dans la littérature est sans doute le quantificateur vectoriel divisé SVQ (Split Vector Quantizer), qui a été d'abord développé par Paliwal et Atal [5] pour les codeurs de parole en bande étroite et

exploré par la suite dans le codage de la parole en large-bande [6]. Dans [7], Biundo et al. ont utilisé un VQ multi-étages divisé (S-MSVQ) avec prédicteur MA de 1^{er} ordre pour le codage des ISFs en large bande. Le même schéma a été utilisé pour coder les paramètres ISF du AMR-WB [3]. Dans [8], So and Paliwal ont proposé le quantificateur vectoriel à divisions commutées SSVQ (Switched Split Vector Quantizer) et le quantificateur multi-trame à base de GMM (Gaussian mixture model) [9]. Dans [10], Xiaochen et al. ont proposé un algorithme efficace de quantification des paramètres ISF basé sur le modèle GMM où les ISFs sont quantifiés par un VQ réseau de point Gaussien. Dans [11], Sheikhan et Garoucy ont proposé trois schémas hybrides pour la quantification des paramètres ISF en large bande. Le premier schéma est basé sur le SSVQ et le S-MSVQ; le second est basé sur le réseau de neurone SOM (Self Organizing Map) et le troisième est basé sur le S-MSVQ et le réseau de neurone GHSOM (Growing Hierarchical SOM).

Dans ce papier, nous proposons une version à débit variable du quantificateur SSVQ développé pour le codage efficace des paramètres ISF du codeur de parole AMR-WB (G.722.2), selon des suppositions de transmission à travers un canal non bruité. Les résultats de simulation montreront que notre schéma de codage des ISFs du G.722.2, nommé codeur ISF-VR-SSVQ, peut fournir des performances comparables à celles du SSVQ conventionnel tout en diminuant le débit de 1 à 2 bits par trame.

2 Quantificateur vectoriel à divisions commutées

Le quantificateur vectoriel à divisions commutées SSVQ (Switched Split Vector Quantizer) est un schéma de codage hybride conçu à base d'un VQ à commutation combiné avec plusieurs quantificateurs vectoriels divisés [8, 9]. Tout d'abord, rappelons brièvement les principes de base du VQ conventionnel et du SVQ.

Un VQ, de dimension k et de débit de 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 [12]. Le principe de conception d'un VQ consiste à décomposer l'espace des vecteurs de source x de dimension k en L classes disjointes $\{R_0, \dots, R_{L-1}\}$ et associer à chaque classe R_i un vecteur-code unique y_i telle que la distorsion totale moyenne D soit minimisée [12]. Dans le passé, plusieurs méthodes de conception d'un VQ optimal ont été développées. La plus populaire est sans doute l'algorithme LBG [13]. Cet algorithme (noté ici LBG-VQ) est une application itérative des deux conditions d'optimalité telle que la partition et le dictionnaire soient mis à jour itérativement. Il converge vers une solution localement optimale selon le choix du dictionnaire initial. Dans nos conceptions à base de systèmes VQ, nous avons utilisé la méthode d'initialisation de Katsavounidis [14] qui nous a permis d'obtenir les meilleurs optimums locaux.

D'autre part, un SVQ de dimension k et de N parties (noté N -SVQ) est composé de N quantificateurs VQ classiques de tailles et de dimensions plus petites [5]. Son principe de base consiste à partitionner l'ensemble des vecteurs x de dimension k de la base d'apprentissage en N sous-ensembles composés de sous-vecteurs de dimension k_i plus petites (avec $\sum_{i=1}^N k_i = k$). Ensuite, pour chaque partie, le dictionnaire VQ correspondant sera conçu en utilisant le LBG-VQ. Un N -SVQ est donc constitué de N dictionnaires VQ de tailles plus petites $L_i = 2^{R_i k_i}$ (où $L = \prod_{i=1}^N L_i$ et R_i est le débit partiel en bpe).

2.1 Principe de conception du SSVQ

Le principe de base du SSVQ consiste à diviser l'espace des vecteurs de la base d'apprentissage en plusieurs parties disjointes (régions de commutation), où chaque partie est représentée par un SVQ local approprié. La figure 1 présente le schéma bloc du principe de construction du dictionnaire SSVQ.

La première étape consiste à appliquer l'algorithme LBG-VQ sur toute la base d'apprentissage afin de produire m vecteurs-code. L'ensemble de ces vecteurs-code est appelé dictionnaire VQ commutateur Y_m où m représente le nombre de direction de commutation. Ensuite, ce dictionnaire sera utilisé pour partager la base d'apprentissage en m parties suivant le critère du voisin le plus proche. Dans la deuxième étape, chaque partie i ($i = 1, \dots, m$) sera représentée par un N -SVQ $_i$ local correspondant.

2.2 Codage/décodage SSVQ

Le codage SSVQ d'un vecteur de source x passe par deux étapes. Au début, le vecteur x est commuté vers l'une des m directions possibles. Par la suite, ce vecteur sera quantifié par le N -SVQ local correspondant qui a été sélectionné par Y_m . Ainsi, le codeur SSVQ transmet au décodeur un indice i composé de $N + 1$ indices binaires concaténés. Le premier indice i_s ($s = 1, \dots, m$) indique la direction de commutation et les N indices binaires i_n ($n = 1, \dots, N$) restants sont fournis par le

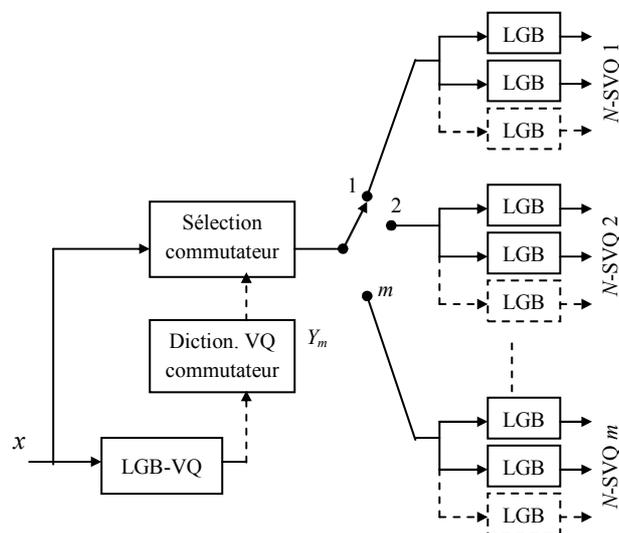


Figure 1 : Construction du dictionnaire SSVQ

N -SVQ $_i$, correspondant à la direction i_s . Pour un SSVQ de b bits/vecteur et m directions de commutation, on requière $b_s = \log_2(m)$ bits pour identifier séparément toute les directions de commutation possibles. Le débit restant de $(b - b_s)$ bits sera partagé en N débits partiels pour coder les N sous-vecteurs de x par le N -SVQ correspondant. Ces débits partiels sont notés par b_j ($j = 1 \dots N$).

Le décodeur SSVQ, qui possède les mêmes dictionnaires que ceux du codeur, reçoit l'indice envoyé $i = (i_s, i_n)$ avec $n = 1, \dots, N$. Il utilise le premier indice i_s pour sélectionner la direction de commutation. Ensuite, il construit le vecteur décodé de x en concaténant les sous vecteurs-code d'indices i_n correspondants au N -SVQ de la partie i_s .

3 Codage efficace des paramètres ISF du codeur AMR-WB

Dans cette section, nous proposons un schéma de codage efficace des paramètres ISF du standard AMR-WB (Rec. G.722.2) pour des transmissions à travers un canal non bruité. Il s'agit d'une version à débit variable du SSVQ conventionnel développée en exploitant la stabilité de l'indice de la direction de commutation pour plusieurs trames successives. Mais d'abords, nous présentons les performances du schéma SSVQ conventionnel appliqué au codage des ISFs du G.722.2.

3.1 Codage des paramètres ISF par SSVQ

Nous présentons, ci-dessous, les performances du codeur des ISFs du G.722.2, conçu à base de la technique SSVQ et appelé "codeur ISF-SSVQ".

Les performances de quantification de nos codeurs des ISFs sont évaluées par la distorsion spectrale SD (Spectral Distorsion). L'expression de la SD pour une trame i est donnée en décibels par [5, 9, 15] :

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

où $S(e^{j2\pi n/N})$ et $\hat{S}(e^{j2\pi n/N})$ représentent respectivement les spectres de puissance original et quantifié du filtre de synthèse LPC de la $i^{\text{ème}}$ trame du signal de parole.

En général, une quantification de qualité transparente est obtenue si les trois conditions suivantes sont maintenues [5]: **1)**- la distorsion spectrale (SD) moyenne est d'environ 1 dB, **2)**- aucune trame externes "outliers" ne doit avoir une SD qui dépasse les 4 dB et **3)**- le pourcentage des trames outliers ayant une SD entre 2 et 4 dB est moins de 2%. Selon Guibé et al. [16] et Cheraitia et Bouzid [15], les tests d'écoute ont montré que ces conditions de quantification transparente, qui sont souvent utilisés dans le cas du codage de la parole en bande étroite, sont aussi valables dans le cas du codage en large-bande.

La base de données parole utilisée dans ce travail se compose d'environ 85 minutes de parole prise de la base de données internationale TIMIT, avec une fréquence

d'échantillonnage de 16 kHz [17]. Pour construire la base des vecteurs ISF, nous avons utilisé la même fonction d'analyse LPC du G.722.2 [3] où une analyse LPC d'ordre 16, par la méthode d'autocorrélation, est effectuée sur chaque trame d'analyse de 20 ms. Une partie de la base de données ISF (208363 vecteurs ISF) est utilisée pour l'apprentissage et la partie restante, de 48606 vecteurs ISF (différente de la base d'apprentissage), est utilisée pour les tests.

Afin d'améliorer davantage les performances de nos codeurs des ISFs et obtenir une quantification transparente à des débits plus bas, nous avons utilisé une mesure de distance euclidienne pondérée plus appropriée [5, 15]:

$$d(f, \hat{f}) = \sum_{i=1}^{16} [w_i (f_i - \hat{f}_i)]^2, \quad (3)$$

où f_i et \hat{f}_i sont respectivement les $i^{\text{ème}}$ coefficients des ISFs original f et quantifié \hat{f} et w_i représente le poids spectral [5] assigné au $i^{\text{ème}}$ coefficient du vecteur ISF.

Pour différents débits de codage b ($b_s + b_1 + \dots + b_5$), les performances d'un exemple de codeur ISF-SSVQ de $m = 16$ directions de commutation sont données dans la Table 1. Notons que dans la conception de notre codeur ISF-SSVQ, les vecteurs ISF large-bande de dimension 16 sont divisés en 5 parties suivant la division (3 – 3 – 3 – 3 – 4) et, dans la mesure du possible, les bits sont uniformément alloués dans chaque régions.

TABLE 1 : Performances des codeurs ISF-SSVQ de 5 parties ($m = 16$) en fonction du débit binaire

Débit - Bits/trame b ($b_s + b_1 + \dots + b_5$)	SD Moy. (dB)	"Outliers" (en %)	
		2-4 dB	> 4 dB
46 (4+9+9+8+8+8)	0.98	0.71	0.00
45 (4+9+8+8+8+8)	1.01	0.86	0.00
44 (4+8+8+8+8+8)	1.03	0.91	0.00
43 (4+8+8+8+8+7)	1.08	1.63	0.00
42 (4+8+8+8+7+7)	1.12	2.12	0.00
41 (4+8+8+7+7+7)	1.16	2.72	0.00

Ces résultats de simulation montrent que le codeur ISF-SSVQ de 5 parties, avec la distance pondérée, peut réaliser une quantification de qualité transparente à 43 bits/trame.

3.2 SSVQ à débit variable pour le codage des ISFs en large bande

Dans cette section, nous proposons une version à débit variable du SSVQ appliqué au codage efficace des paramètres ISF du G.722.2, selon des suppositions de transmission idéales à travers un canal non bruité.

Le nouveau schéma, nommé ISF-VR-SSVQ, garde pratiquement le même concept de construction du SSVQ de base. Le changement se fait au niveau du codage/décodage où l'on exploite la stabilité de l'indice de commutation i_s pour plusieurs trames successives. En effet, une étude statistique de la variation de cet indice

(Figure 2) a montrée que plusieurs vecteurs ISF de trames successives peuvent être codés avec le même quantificateur local N -SVQ de direction i_s . Cette observation nous a donnée l'idée de concevoir un SSVQ à débit variable en ajoutant un bit supplémentaire à l'ensemble des bits transmis afin d'indiquer au décodeur de maintenir l'indice i_s précédent (sans le transmettre) ou d'utiliser un autre indice qui sera transmis. Ainsi, si la direction de commutation actuelle est identique à celle de la trame précédente, le codeur SSVQ transmet seulement un bit supplémentaire "0" sans transmettre l'indice i_s . Dans ce cas, une diminution du débit de $b_s - 1$ bits/trame peut être obtenue lors de la transmission comparé au codeur ISF-SSVQ conventionnel.

Dans le cas contraire, le codeur transmet un bit supplémentaire "1" en plus du nouveau indice i_s ; ce qui laisse penser que cette idée peut causer une augmentation du débit de codage. Ceci n'étant pas vrai puisque les résultats de simulation, présentés ci-dessous, montreront que le codeur ISF-VR-SSVQ peut assurer des performances comparables à celles du ISF-SSVQ avec une réduction du débit de codage.

Les performances de notre codeur ISF-VR-SSVQ de 5 parties sont présentées dans la Table 2. Notons que les codeurs ISF-VR-SSVQ de 5 parties ont été conçus dans les mêmes conditions que les codeurs ISF-SSVQ de 5 parties (c-à-d., $m = 16$, division, allocation de bits, base de données ISF, distance pondérée). Cependant, le débit de codage réel est déterminé à la fin de la transmission car il dépend de la stabilité de la valeur de l'indice de commutation d'une trame vers une autre, comme expliqué précédemment.

TABLE 2 : Performances des codeurs ISF-VR-SSVQ de 5 parties ($m = 16$) en fonction du débit variable

Débit Bits/trame	SD Moy. (dB)	"Outliers" (en %)	
		2-4 dB	> 4 dB
44.79	0.98	0.71	0.00
43.79	1.01	0.86	0.00
42.79	1.03	0.91	0.00
41.79	1.08	1.63	0.00

En comparant ces résultats avec ceux données dans la Tables 1, on remarque clairement que le codeur ISF-VR-SSVQ assure pratiquement les mêmes performances que ceux du codeur ISF-SSVQ mais avec une diminution du débit de codage. En effet, le codeur ISF-VR-SSVQ de 5 parties a besoin de seulement de 41.79 bits/trame pour réaliser une quantification de qualité transparente

4 Conclusion

Dans ce travail, une version à débit variable du SSVQ a été développée pour le codage efficace des paramètres ISF en large bande du codeur G.722.2, selon des suppositions de transmission à travers un canal non bruité. Les résultats de simulation ont montré que notre codeur ISF-VR-SSVQ fournit des performances

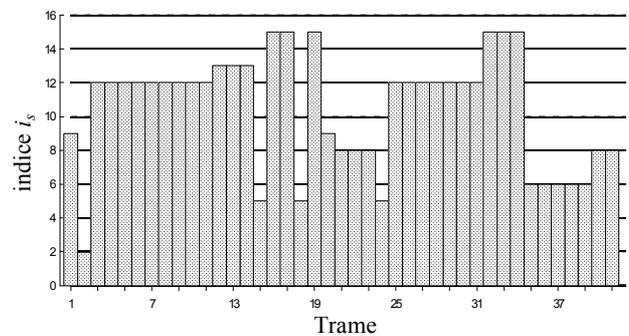


Figure 2: Exemple de variation de l'indice de commutation durant le codage SSVQ

comparables à celles du ISF-SSVQ conventionnel tout en assurant un gain de 1-2 bits/trame.

Les performances du codeur ISF-VR-SSVQ en présence des erreurs de canal reste à être étudié.

Références

- [1] W. B. Kleijn, and K. K. Paliwal, *Speech coding and synthesis*, Elsevier Science B.V., 1995.
- [2] L. R. Rabiner, and R. W. Schafer, *Digital Processing of speech signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.
- [3] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620-636, November 2002.
- [4] Y. Bistriz, H. Lev-Ari, and T. Kailath, "Immittance domain Levinson algorithms," *IEEE transactions on information theory*, vol. 35, pp. 675-682, 1989.
- [5] 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, January 1993.
- [6] J. H. Chen, and D. Wang, "Transform predictive coding of wideband speech signals," *Proc. of the ICASSP'96*, Atlanta, USA, May 1996, pp. 275-278.
- [7] G. Biundo, S. Grassi, M. Ansoorge, F. Pellandini, and P. A. Farine, "Design techniques for spectral quantization in wideband speech coding," *3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication, Budapest, Proceedings*, 2002, pp. 114-119.
- [8] S. So and K. K. Paliwal, "Efficient vector quantization of line spectral frequencies using the switched split vector quantiser," *Proc. of Int. Conf. Spoken language Processing*, Jeju, Korea, 2004.
- [9] S. So and K. K. Paliwal, "comparative study of LPC parameter representations and quantisation schemes for wideband speech coding," *Digital Signal Processing, Elsevier*, vol. 17, pp. 114-137, 2007.
- [10] W. Xiaochen, Z. Yong, H. Ruimin, and D. Xi, "An Immittance Spectral Frequency parameters quantization Algorithm based on Gaussian Mixture Model," *International Conference on Multimedia Information Networking and Security, MINES'09, Proceedings*, 2009, pp. 324-328.
- [11] M. Sheikhan, and S. Garoucy, "Hybrid VQ and neural models for ISF quantization in wideband speech coding," *World Applied Sciences Journal*, vol. 10, pp. 59-66, 2010.
- [12] A. Gersho and R. M. Gray, *Vector quantization and Signal compression*, Kluwer Academic Publishers, USA, 1992.
- [13] Y. Linde, A. Buzo and R.M. Gray, "An Algorithm for Vector Quantization Design," *IEEE Transactions on Communications*, vol. COM-28, pp. 84-95, 1980.
- [14] I. Katsavounidis C. Kuo, and Z. Zhang, "A new initialization technique for generalized Lloyd iteration," *IEEE Signal Proc. Letters*, vol. 1, pp. 144-146, October 1994.
- [15] S. Cheraitia and M. Bouzid, "Robust coding of wideband speech immittance spectral frequencies," *Speech Communication, Elsevier*, vol. 65, pp. 94-108, Nov. 2014.
- [16] G. Guibé, H. T. How, L. Hanzo, "Speech spectral quantizers for wideband speech coding," *European Transactions on Telecommunications*, vol. 12, no. 6, pp. 535-545, 2001.
- [17] *DARPA TIMIT Acoustic-phonetic Continuous Speech Database*, Technology Building, National Institute of Standards and Technology (NIST), Gaithersburg October 1988.