

Transcodage intelligent à faible complexité entre les codeurs UIT-T G.729 (8 kbit/s) et 3GPP NB-AMR (12.2 kbit/s)

Mohamed Ghenania¹ Claude Lamblin¹

¹ France Telecom R&D – DIH/IPS 2 avenue Pierre Marzin, 22307 Lannion Cedex, FRANCE

{mohamed.ghenania, claude.lamblin}@francetelecom.com

Résumé

L'interconnexion des réseaux pose le problème de l'interopérabilité entre les différents standards de codage de la parole. Généralement, dans les chaînes de communications, un transcodage en tandem (décodage/recodage) est utilisé. Néanmoins, des solutions de transcodage intelligent exploitant les similitudes entre formats sont proposées aujourd'hui. Elles permettent de s'affranchir des inconvénients liés au tandem (complexité, retard et dégradation du signal). Cet article en dresse tout d'abord un panorama pour les codeurs CELP. Puis un algorithme de transcodage à faible complexité entre les codeurs UIT-T G.729 (8 kbit/s) et 3GPP NB-AMR (12.2 kbit/s) est présenté. Il comporte quatre modules correspondant aux conversions des quatre paramètres CELP: coefficients LSP, excitations adaptative et fixe et gains associés. Une méthode originale de conversion est proposée pour l'excitation fixe. Enfin, des tests de qualité montrent que notre solution conduit à une qualité très proche du tandem pour une complexité bien inférieure et un retard algorithmique réduit.

Mots clefs

Codage de la parole, transcodage, ACELP.

1 Introduction

Ces dernières années, la demande en communications de qualité a considérablement augmenté parallèlement à la multiplicité des terminaux et des accès. Pour fournir un accès multimédia universel à l'utilisateur, des réseaux de communication utilisant des standards de codage de la parole différents doivent être interconnectés. Bien que basés sur le même principe de codage CELP (Code-Excited Linear Prediction), ces différents standards sont incompatibles entre eux. Pour résoudre ce problème d'interopérabilité, la solution la plus simple consiste à décoder chaque trame codée au premier format pour ensuite la recoder au second format. Cependant, cette méthode appelée transcodage en tandem présente trois inconvénients majeurs: une complexité de calcul importante, l'introduction d'un délai algorithmique supplémentaire dans la communication, et une dégradation du signal. Pour s'affranchir de ces inconvénients des solutions de transcodages dits "intelligents", ont été récemment proposées. Elles exploitent les similitudes entre formats de codage CELP à partir d'une approche par

paramètres. Cet article dresse tout d'abord (§2) un panorama de ces techniques de transcodages intelligents entre standards de codage CELP. Puis un nouvel algorithme de transcodage intelligent entre les codeurs UIT-T G729 (à 8 kbit/s) [1] et 3GPP NB-AMR (à 12.2kbits/s) [2] est présenté (§3). En particulier, une nouvelle méthode de transcodage à faible complexité est proposée pour les répertoires d'excitation fixe ACELP. Enfin, la section 4 est consacrée à la comparaison des performances du tandem et de la solution proposée.

2 Panorama des techniques de transcodage CELP

La modification du format de codage, ou transcodage, lors d'une communication est un problème récurrent aujourd'hui. Plusieurs solutions ont été proposées. La plus élémentaire (et la plus courante à l'heure actuelle), appelée tandem, est la mise bout à bout d'un décodeur et d'un codeur. La trame codée sous un premier format est décodée. Le signal synthétisé est alors recodé sous un second format accepté par la suite de la chaîne de communication. Cette solution est très coûteuse en complexité (essentiellement à cause du recodage), elle introduit un retard supplémentaire, et la dégradation par rapport au signal original est amplifiée par le second codage car ce dernier s'effectue sur un signal décodé.

Toutefois, plusieurs solutions de transcodage dit intelligent existent aujourd'hui. En exploitant les similitudes entre formats, elles réduisent la complexité et le retard algorithmique du transcodage par rapport au tandem pour une qualité sensiblement équivalente. Tous les codeurs CELP actuels calculent, quantifient et transmettent les mêmes paramètres: coefficients LSP, excitation adaptative (EA), excitation fixe (EF) et gains d'excitation associés. Pour un même paramètre, les différences entre codeurs résident dans la méthode de calcul, la fréquence de calcul ou la quantification. Dès lors, trois cas sont possibles pour transcoder un paramètre. Si un paramètre est calculé et quantifié de la même façon par les deux formats de codage A et B, son transcodage se fait au niveau binaire par recopie de son champ binaire du train binaire de A dans celui de B. C'est le cas pour l'excitation adaptative dans [3, 6-7] et fixe dans [6-7].

Si le paramètre est calculé de la même manière mais quantifié différemment, il faut le requantifier avec la méthode utilisée par le format B. De même, si A et B ne le

calculent pas à la même fréquence (par exemple si leurs longueurs de sous-trame sont différentes), il faut interpoler (ou décimer) ce paramètre. Ceci peut être réalisé en décodant seulement le paramètre sans avoir à remonter jusqu'au niveau signal. Le transcodage est réalisé au niveau paramètre. Les coefficients LSP sont généralement transcodés à ce niveau [3-10]. Le transcodage au niveau paramètre est aussi utilisé pour les gains d'excitations adaptative [6-7] et fixe [6-7, 10].

Enfin, si les traitements par deux formats d'un paramètre sont trop différents, il devient nécessaire de décoder le signal et de calculer ce paramètre comme dans le cas du tandem. On parle alors de transcodage au niveau signal. Cette méthode est souvent employée pour la conversion de l'excitation fixe [4, 5, 8, 9], mais aussi pour les gains d'excitations adaptative [3, 9, 10] et fixe [3, 4, 5, 8, 9]. De manière générale, le transcodage intelligent a pour objectif de remonter le moins possible au niveau signal, ce que fait systématiquement le tandem. Si cela est indispensable il existe des alternatives intelligentes: dans certains cas, il suffit de ne remonter qu'au niveau de l'excitation (gains dans [7]). Par ailleurs, des solutions ont été développées pour réduire la complexité d'une conversion au niveau signal en utilisant les informations issues du premier format relatives au paramètre à transcoder. Par exemple, en restreignant la plage de valeurs testées par la boucle ouverte d'estimation du pitch du format B autour de la valeur de pitch du format A [4-5, 8-10]. De la même manière, l'exploration du dictionnaire de gains multitap du G.723.1 peut être limitée en fonction du gain monotap choisi par le premier format [4-5, 8]. Enfin, des solutions analogues ont aussi été proposées afin de limiter l'exploration des répertoires d'excitation fixe [3, 10].

3 Méthode de transcodage entre l'UIT-T G.729 et le 3GPP NB-AMR

Ces trois niveaux de conversion (binaire, paramètre et signal) ont été utilisés pour développer un algorithme de transcodage intelligent performant entre les codeurs UIT-T G.729 à 8kbit/s et 3GPP NB-AMR à 12.2 kbit/s. Basés sur la technologie ACELP (Algebraic Code-Excited Linear Prediction), ces deux standards sont des codeurs multi-débits. Dans cet article, seuls sont considérés le mode principal du G.729 et le mode 12.2 kbit/s du NB-AMR, aussi appelé EFR (Enhanced Full Rate), car ce mode est identique au codeur GSM-EFR.

	G.729	EFR
Sous-trames	2×5ms	4×5ms
LSP	1 QV pour 10ms	1 SMQ pour 20ms
E. A.	Monotap 1/3	Monotap 1/6
E. F.	4 impulsions	10 impulsions
gain E.A.	QV sur 7bits	QS (4bits)
gain E.F.		QS (5bits)

Tableau 1: Similitude et différences entre G.729 et EFR

Les deux codeurs ont des similitudes et des différences résumées dans le tableau 1. Pour les trois premiers paramètres CELP, notre solution propose des techniques de transcodage intelligent exploitant les similitudes entre ces deux formats. En particulier, une méthode originale est proposée pour transcoder l'excitation fixe.

3.1 Conversion des coefficients LSP

Les deux codeurs effectuent une analyse LPC toutes les 10 ms (i.e. une par trame pour G.729 et 2 par trame pour EFR) puis convertissent les 10 coefficients LPC obtenus en coefficients LSP avant de les quantifier au moyen d'une quantification vectorielle prédictive à moyenne ajustée. Néanmoins, G.729 utilise pour un jeu de LSP une quantification vectorielle prédictive commutée d'ordre 4 sur 18 bits. De plus, son analyse LPC introduit un *look-ahead* de 5ms. L'EFR, quantifie conjointement deux jeux de LSP par trame sur 38 bits par une quantification matricielle par code-produit (*Split Matrix Quantization*) à moyenne ajustée d'ordre 1 appliquée sur les LSF résiduels. Ici, le transcodage se fait donc au niveau paramètre: les coefficients LSP du format A sont décodés puis re-quantifiés par la méthode du format B. Avant quantification, la correspondance entre les LSP de l'EFR et du G.729 est donnée par les relations suivantes:

EFR → G.729:

$$LSP_{2n}^{G729} = LSP_{n-1}^{EFR} \text{ et } LSP_{2n+1}^{G729} = LSP_{n-3}^{EFR}$$

G.729 → EFR:

$$LSP_{n-1}^{EFR} = LSP_{2n}^{G729} \text{ et } LSP_{n-3}^{EFR} = LSP_{2n+1}^{G729}$$

où LSP_n^{G729} (resp. LSP_{n-k}^{EFR}) est le jeu de LSP de la trame n du G.729 (resp. de la $k^{\text{ème}}$ sous-trame, $k \in [0,3]$, de la trame n de l'EFR).

La complexité de calcul est fortement réduite, car l'analyse LPC est éliminée ainsi que la conversion LPC vers LSP. Le retard algorithmique est aussi réduit de 5ms (suppression du look-ahead du G.729). De plus, la distorsion spectrale apportée par la conversion est plus faible que celle apportée par le tandem. Le tableau 2 compare les distorsions spectrales moyennes (calculées sur 18 phrases en français avec 3 locuteurs) du transcodeur et du tandem. Les LSP transcodés sont plus proches des originaux (i.e. quantifiés une seule fois) que ceux obtenus par le tandem. En effet, celui-ci quantifie et transmet des LSP calculés sur un signal déjà codé-décodé par A alors que le transcodeur re-quantifie des LSP calculés sur le signal original et juste quantifiés par A.

DS (dB)	G.729 vers EFR			EFR vers G.729		
	Moy.	Outliers (2-4dB)	Outliers (>4dB)	Moy.	Outliers (2-4dB)	Outliers (>4dB)
Tandem	2.71	52.38%	5.40%	2.29	32.59%	1.01%
Transcodeur	1.08	0.52%	0%	1.14	1.18%	0%

Tableau 2: distorsion spectrale tandem/transcodeur

La figure 1 (transcodage G729 vers EFR) montre pour une trame voisée que le spectre LPC obtenu est plus proche du spectre original que ne l'est celui du tandem.

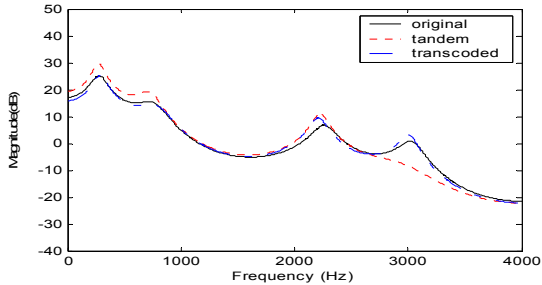


Figure 1: Comparaison des spectres LPC

3.2 Conversion du pitch

Le G.729 et l'EFR estiment le délai de l'excitation adaptative en combinant une recherche en boucle ouverte avec une recherche en boucle fermée. Le délai obtenu (appelé *pitch*) est codé en valeur absolue - avec une résolution fractionnaire s'il est inférieur à une valeur limite pour les sous-trames paires. Il est codé en valeur différentielle (différence avec le délai précédent) pour les sous-trames impaires avec une résolution fractionnaire. Toutefois, comme le montre le tableau 3, les modélisations de pitch des deux codeurs comportent des différences: l'EFR a une résolution fractionnaire (1/6) plus fine que le G.729 (1/3). La plage de valeurs de pitch de l'EFR [17_{3/6} 143] est plus grande que celle du G.729 [19_{1/3} 143]. La valeur limite entre résolution fractionnaire et entière est égale à 95 pour l'EFR et 85 pour le G.729.

sous-trames	paires			impaires	
	Fractionnaire		Entier	Fractionnaire	
	frac	Plage		frac	Plage
G.729	1/3	19 _{1/3} 84 _{2/3}	85 143	1/3	-5 _{2/3} 4 _{2/3}
EFR	1/6	17 _{3/6} 94 _{3/6}	95 143	1/6	-5 _{3/6} 4 _{3/6}

Tableau 3: modèles de pitch du G.729 et de l'EFR

En exploitant l'inclusion de la plage de valeurs du G729 dans celle de l'EFR, la conversion du pitch peut se faire au niveau paramètre ou au niveau binaire, en arrondissant la fraction si nécessaire pour le transcodage EFR vers G.729. On élimine ainsi les deux recherches coûteuses en boucle ouverte et boucle fermée sans perte sensible de qualité.

3.3 Conversion de l'excitation fixe

L'EFR et le G.729 utilisent pour leur excitation fixe un modèle ACELP (tableaux 4 et 5). L'EFR a un répertoire ACELP à 35 bits pour 10 impulsions d'amplitude +/-1 et le G729 un répertoire ACELP à 17 bits pour 4 impulsions d'amplitudes +/-1. Une méthode originale et à faible complexité est proposée ici pour le transcodage de ce paramètre. Ce transcodage est réalisé au niveau signal, mais le nombre de combinaisons de positions testées est fortement réduit par rapport au tandem. En effet, en

restreignant l'ensemble des positions testées aux positions du codeur A et à leurs voisinages, on n'explore qu'un sous-ensemble réduit du répertoire ACELP du codeur B.

Piste	Impulsion	Signe	Positions
T ₀	i ₀ , i ₅	± 1	0, 5, 10, 15, 20, 25, 30, 35
T ₁	i ₁ , i ₆	± 1	1, 6, 11, 16, 21, 26, 31, 36
T ₂	i ₂ , i ₇	± 1	2, 7, 12, 17, 22, 27, 32, 37
T ₃	i ₃ , i ₈	± 1	3, 8, 13, 18, 23, 28, 33, 38
T ₄	i ₄ , i ₉	± 1	4, 9, 14, 19, 24, 29, 34, 39

Tableau 4: répertoire ACELP de l'EFR

Piste	Impulsion	Signe	Positions
T ₀	i ₀	± 1	0, 5, 10, 15, 20, 25, 30, 35
T ₁	i ₁	± 1	1, 6, 11, 16, 21, 26, 31, 36
T ₂	i ₂	± 1	2, 7, 12, 17, 22, 27, 32, 37
T ₃ T ₄	i ₃	± 1	3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39

Tableau 5: répertoire ACELP du G.729 (8 kbit/s)

De plus, ces positions privilégiées sont rapidement extraites en exploitant la structure en pistes entrelacées ISSP (Interleaved Single-Pulse Permutation) commune aux deux répertoires. Les deux codeurs découpent la sous-trame de 40 positions en 5 pistes entrelacées de longueur 8. L'EFR place deux impulsions par pistes. G.729 en place une sur les trois premières pistes et une sur la quatrième ou la cinquième piste.

Chaque position d'impulsion peut être représentée par le couple (t, r) où t est la piste occupée et r son rang dans la piste: $p = 5r + t$ avec $0 \leq t \leq 4$ et $0 \leq r \leq 7$.

Pour la position de chaque impulsion du format A, les voisinages droit et gauche sont extraits comme suit: Soit d la taille du voisinage choisie, la position voisine $p_k^{(i)}$ de la position $p^{(i)} = (t_j, r)$ de la i^{ème} impulsion est:

$$p_k^{(i)} = p^{(i)} + k = (t_k^{(i)}, r_k^{(i)}) \text{ avec } k \text{ dans } [-d, d],$$

$$\text{où } t_k^{(i)} = t_{(j+k) \bmod 5}^{(i)}, \text{ et } r_k^{(i)} = r + \left\lfloor \frac{j+k}{5} \right\rfloor$$

A noter que $0 \leq r \leq 7$, et que $p_0^{(i)} = p^{(i)}$.

Pour chaque piste T_j ($0 \leq j \leq 4$), les positions autorisées doivent appartenir à l'union des voisinages des N

$$\text{impulsions du codeur A: } S = \bigcup_{i=1}^N \left\{ \bigcup_{k=-d}^d p_k^{(i)} \right\}$$

Cette focalisation conduit à n'explorer qu'un sous-ensemble réduit du répertoire du codeur B et réduit fortement la complexité sans perte sensible de qualité. Le

tableau 6 montre pour les deux sens de transcoding, que le critère CELP CC_{foc} obtenu avec une exploration restreinte est proche du critère CELP standard CC_{std} obtenue normalement par le codeur B. Des tests informels de comparaison par paires faits par des experts indiquent que cette méthode n'introduit pas de dégradation audible.

	d	$CC_{foc} = CC_{std}$	$CC_{foc} > 0.8 CC_{std}$
G.729 → EFR	2	12.4%	75.2%
EFR → G.729	1	60.1%	87.4%

Table 6: comparaison des critères CELP

3.4 Conversion des gains

Le G729 quantifie conjointement les deux gains d'excitation (adaptative et fixe) sur 7 bits par une quantification vectorielle. L'EFR, quant à lui, quantifie les deux gains séparément par une quantification scalaire respectivement sur 4 et 5 bits. Ces deux paramètres sont recalculés puis quantifiés comme le fait le tandem.

4 Évaluation des performances

4.1 Complexité de calcul

Pour les 4 paramètres, le tableau 7 compare le gain relatif en complexité de notre algorithme par rapport au tandem.

paramètre	G.729 → EFR	EFR → G.729
LPC	50%	40%
E. A.	99%	99%
E. F.	38%	43%
Gains	0%	0%

Tableau 7: Complexité de calcul des paramètres CELP

4.2 Qualité

La comparaison des qualités restituées par le transcodeur et le tandem mesurées objectivement (PESQ sur 12 phrases -6 hommes et 6 femmes- en français) indique que ces qualités sont sensiblement équivalentes (tableau 8).

	G.729 → EFR		EFR → G.729	
	Homme	Femme	Homme	Femme
Tandem	3.650	3.346	3.716	3.448
Transcodeur	3.619	3.225	3.632	3.256

Tableau 8: Comparaison des scores PESQ

Des tests subjectifs de comparaison par paires sur combiné ont été réalisés par 16 auditeurs sur 16 double phrases de 8 locuteurs (3 hommes, 3 femmes, 2 enfants). Dans la moitié des cas, les deux solutions sont indifférenciées. Toutefois, le tandem a plus souvent été préféré à notre solution, en particulier dans le cas EFR vers G.729. Les dégradations apportées par les deux méthodes ont été perçues différemment: le tandem introduit une distorsion spectrale (relativement masquée par le combiné) alors que le transcodeur accroche localement.

	Tandem	pas de préférence	Transcodeur
G.729 → EFR	28.5%	53.5%	18%
EFR → G.729	37.5%	50.8%	11.7%

Tableau 9: Tests de comparaison par paires

4.3 Retard

Le retard introduit est réduit grâce notamment à un temps de calcul plus faible, et l'élimination du délai algorithmique introduit par l'analyse LPC.

5 Conclusion

Dans cet article, après avoir dressé un panorama des techniques de transcoding entre codeurs CELP, un transcodeur intelligent à faible complexité entre les codeurs UIT-T G.729 et 3GPP NB-AMR a été proposé. Il met en œuvre une solution originale de conversion de l'excitation fixe. Cette solution a aussi été étendue à d'autres transcodages entre d'autres modèles multi-impulsionnels d'excitation fixe comme ceux des codeurs UIT-T G.723.1 ou les autres modes du codeur 3GPP NB-AMR. Là encore, elle a permis une forte réduction de la complexité pour une qualité proche du tandem.

Références

- [1] IUT-T Rec. G.729, "Coding of speech at 8 kbit/s using conjugate structure algebraic-Code-Excited Linear Prediction (CS-ACELP)", Mars 1996.
- [2] 3GPP, "AMR speech codec: Transcoding functions", December 1999, 3G TS 26.090.
- [3] Shu-Min Tsai, Jar-Ferr Yang, "GSM to G.729 speech transcoder", in *Proc. ICECS 2001*, pp. 485-488.
- [4] Sung Wan Yoon and al., "An efficient Transcoding Algorithm for G723.1 and G.729A speech coders", in *Proc. Eurospeech 2001*, pp. 2499-2502.
- [5] Kyung Tae Kim and al., "An efficient transcoding algorithm for G723.1 and EVRC speech coders", in *Proc. IEEE VTS 2001*, pp.1561-1564.
- [6] Ysujji Ota and al., "Speech Coding Translation for IP and 3G Mobile Integrated Network", in *Proc. ICC 2002*, pp.114-118.
- [7] H.G. Kang and al., "Improving transcoding capability of speech coders in clean and frame erased channel environments", *Proc. IEEE Workshop on Speech Coding 2000*, pp78-80.
- [8] Sung-Wan Yoon and al., "Transcoding algorithm for G723.1 and AMR speech Coders: For interoperability between VoIP and Mobile Networks", *Proc. Eurospeech 2003*, pp. 1101-1104.
- [9] Seongho Seo and al., "A novel Transcoding Algorithm for SMV and G723.1 speech coders via direct parameter Transformation", in *Proc. Eurospeech 2003*, pp.2861-2864.
- [10] Seongho Seo and al., "A novel Transcoding Algorithm for AMR and EVRC speech codecs via direct parameter Transformation", in *Proc. ICASSP 2003*, pp. 177-180, vol. II.