

Contrôle de l’audibilité du bruit de quantification induit par la pré-distorsion d’un signal de parole

Gaël MAHÉ¹, André GILLOIRE²

¹Université René Descartes - Paris 5 / CRIP5
45 rue des Saints-Pères, 75270 Paris Cedex 06, France

²France Télécom R&D / DIH / IPS
2 avenue Pierre Marzin, 22307 Lannion Cedex, France

mahe@math-info.univ-paris5.fr, andre.gilloire@rd.francetelecom.com

Résumé – La conjugaison de la prédistorsion et de la quantification d’un signal de parole peut conduire, après ré-équilibre spectral du signal, à une dégradation du rapport signal à bruit en réception. Nous proposons donc, dans ce contexte, deux méthodes de reformage spectral du bruit de quantification visant à masquer celui-ci, l’une fondée sur la réinjection, à l’entrée du quantificateur, de l’erreur de quantification filtrée, l’autre consistant à rechercher la séquence quantifiée maximisant un critère probabiliste de masquage. Les résultats objectifs montrent une forte capacité de masquage, tandis que l’évaluation subjective témoigne de la gêne perceptive occasionnée par le bruit résiduel sporadiquement non masqué.

Abstract – The combination of predistorsion and quantization of a speech signal can lead to a decrease of the signal to noise ratio after inversion of the pre-distortion. In this context, we propose two noise spectral shaping methods aiming at masking the quantization noise, the first one based on the feedback of the quantization error to the input of the quantizer, the other one consisting in searching for the quantized sequence that maximizes a probabilistic criterion of masking. The objective results show that the noise is mostly masked, whereas a subjective evaluation indicates that the residual non masked noise is perceptually disturbing.

1 Introduction

Lorsqu’un signal en entrée d’une chaîne de communication est prédistoré spectralement de manière à compenser par anticipation le filtrage de cette chaîne, il peut en résulter un fort déséquilibre entre les composantes spectrales du signal. Si une opération de quantification sans mémoire (de type MIC) intervient après la pré-distorsion, il apparaît que le niveau du bruit de quantification est proche de celui des composantes les plus faibles. La prédistorsion provoque alors, après le ré-équilibre du spectre du signal, une dégradation du rapport signal à bruit en réception.

Un exemple de ce phénomène est donné par l’égaleur centralisé proposé dans [4], visant à corriger les dégradations du timbre de la parole téléphonique. Ces dégradations proviennent des distorsions spectrales introduites lors de la transmission de la parole en bande de base analogique : filtrage passe-bande des terminaux en émission et en réception [7], et filtrage passe-bas des lignes analogiques locales. L’égaleur est placé au cœur du réseau numérique, comme indiqué sur la figure 1, et compense ces filtrages sur une bande limitée, typiquement 200 Hz à 3150 Hz. De manière à anticiper l’atténuation des composantes basses-fréquences (200-400 Hz) par le système de réception, celles-ci sont sur-amplifiées par l’égaleur.

Le signal de parole en réception de la liaison ainsi égalisée est affecté d’un bruit de quantification très perceptible, de niveau variable selon les phonèmes. Observons ce phénomène pour le phonème [Y] prononcé par une locutrice féminine. La figure 2 représente les enveloppes spectrales du signal et des bruits de quantification en trois points successifs de la liaison :

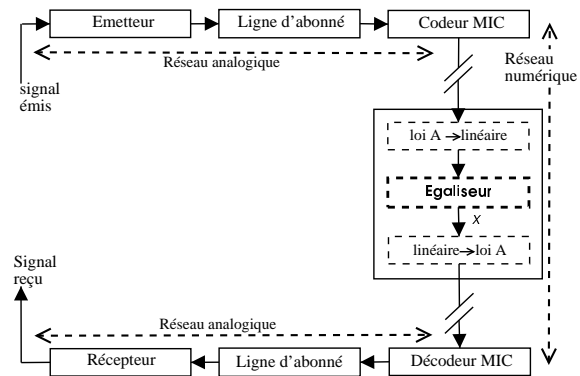


FIG. 1 – Compensation des distorsions spectrales d’une liaison téléphonique par un égaliseur centralisé

après le codage MIC ; après l’égaleur le cas échéant ; en réception avec ou sans égaliseur. Les bruits du codeur MIC et du convertisseur linéaire-loi A sont notés respectivement q_0 et q_1 . Le phonème [Y] possède un premier formant vers 200 Hz, qui est fortement atténué par le système d’émission. Ainsi, la bande 200-300 Hz, après sur-amplification par l’égaleur, est fortement énergétique et détermine le niveau du bruit de quantification du convertisseur linéaire-loi A, qui dépasse alors celui du deuxième formant (vers 2500 Hz), contrairement à q_0 .

Nous proposons dans la section 2 deux méthodes de reformage spectral du bruit de quantification visant à masquer celui-ci. Ces méthodes sont évaluées selon des critères objectifs dans la section 3. La section 4 présente l’évaluation subjective formelle d’une de ces méthodes.

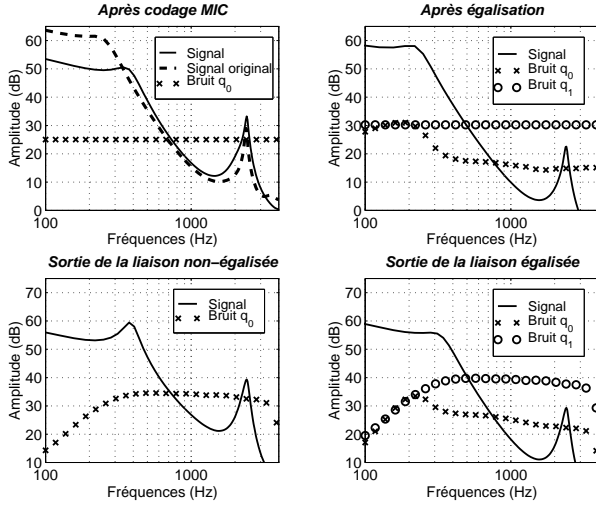


FIG. 2 – Enveloppe spectrale de $[Y]$ et des bruits de quantification correspondants en différents points d'une liaison

2 Méthodes de reformage spectral du bruit

Afin de masquer le bruit de quantification, nous proposons d'agir sur l'opération de quantification qui suit l'égaliseur, de telle sorte que le spectre du bruit soit sous la courbe de masquage du signal :

$$\gamma_b(f) = \lambda^2 \gamma_{\text{Masque}}(f) \quad (1)$$

avec γ_b la densité spectrale de bruit, γ_{Masque} le seuil de masquage et λ un facteur inférieur à 1. Les méthodes usuelles de reformage spectral du bruit d'un quantificateur instantané [6, 1] consistent à blanchir le signal à quantifier et reformer son spectre en réception. Nous proposons ici deux méthodes [5] permettant de reformer le bruit uniquement en modifiant la quantification.

2.1 Reformage par réinjection de l'erreur de quantification

La première méthode, inspirée de [6], consiste à réinjecter à l'entrée du quantificateur l'erreur de quantification filtrée, tel que représenté sur la figure 3. D'après cette structure, la transformée en Z du bruit b s'exprime par :

$$B(z) = \tilde{X}(z) - X(z) = (1 + G(z))Q(z) \quad (2)$$

Or le reformage spectral du bruit implique, selon (1) :

$$B(z) = \lambda H(z)W(z) = \lambda \sigma (1 + G_0(z))W(z) \quad (3)$$

où $H(z) = \sigma(1 + G_0(z))$ est un modèle ARMA du masque et w un bruit blanc de variance 1. On en déduit le filtre de boucle G et l'expression de λ :

$$G(z) = G_0(z), \lambda = \frac{\sigma q}{\sigma} \quad (4)$$

où σ_q est l'écart type de l'erreur de quantification q .

2.2 Approche probabiliste

Le masquage du bruit implique la minimisation de λ , donc la maximisation de σ . Ce problème est fortement contraint par

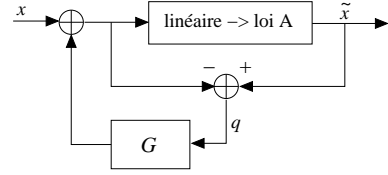


FIG. 3 – Reformage spectral du bruit de quantification par réinjection de l'erreur de quantification

la nécessité de stabilité de la boucle de rétroaction. Pour s'affranchir de cette contrainte, nous proposons une seconde méthode, fondée sur une approche probabiliste : au lieu de quantifier chaque échantillon par la valeur de quantification la plus proche, nous quantifions globalement une séquence de signal en recherchant le chemin de quantification (séquence d'échantillons quantifiés) de 0 à n , noté $C(0..n)$, le plus probable, connaissant le spectre que doit avoir le bruit de quantification. Les chemins de quantification possibles sont construits pas à pas, en utilisant la relation suivante :

$$\Pr(C(0..n)) = \Pr(Q(n)|C(0..n-1)) \Pr(C(0..n-1)) \quad (5)$$

où $Q(n)$ est la valeur de quantification associée à l'échantillon égalisé (pré-distordu) $x(n)$. Si nous considérons la quantification comme l'ajout d'un bruit b ayant le spectre souhaité suivi d'une comparaison à un seuil, la probabilité conditionnelle de quantifier $x(n)$ par une valeur quantifiée Q_k est définie par :

$$\begin{aligned} \Pr(Q(n) = Q_k | C(0..n-1), \gamma_b) \\ = \Pr(S_k < x(n) + b(n) < S_{k+1} | C(0..n-1), \gamma_b) \end{aligned} \quad (6)$$

où S_k et S_{k+1} sont les seuils inférieur et supérieur associés à Q_k .

Nous approchons la courbe de masquage par la densité spectrale de puissance associée à un modèle ARMA $\{(a_i)_{1 \leq i \leq p}; (d_j)_{1 \leq j \leq q}; \sigma\}$ qui est celui de b à un facteur positif λ_0 près, fixé à une valeur inférieure à 1 :

$$\begin{aligned} b(n) = \lambda_0 \sigma' w(n) + x(n) \\ - \sum_{i=1}^p a_i b(n-i) + \sum_{j=1}^q d_j \lambda_0 \sigma' w(n-j) \end{aligned} \quad (7)$$

où w est un bruit blanc de variance unité. Ainsi, à chaque instant n , $x(n) + b(n)$ est une variable aléatoire de même distribution que $w(n)$ et d'espérance :

$$\begin{aligned} E[x(n) + b(n)] = x(n) - \sum_{i=1}^p a_i b(n-i) \\ + \sum_{j=1}^q d_j \lambda_0 \sigma' w(n-j) \end{aligned} \quad (8)$$

Suivant la distribution de $w(n)$ choisie, nous pouvons alors calculer la probabilité de tout chemin de quantification.

Les chemins les plus probables sont sélectionnés selon un algorithme de type Viterbi. Pour chaque suite $C(n+1..N)$ d'un chemin $C(0..n)$, avec N le nombre total d'échantillons à quantifier,

$$\begin{aligned} \Pr(C(0..n) \circ C(n+1..N)) \\ = \Pr(C(0..n)) \Pr(C(n+1..N)|C(0..n)) \end{aligned} \quad (9)$$

où \circ désigne la concaténation. D'après ce qui précède, la probabilité $\Pr(C(n+1\dots N)|C(0\dots n))$ ne dépend, à l'instant n que des échantillons de x et Q postérieurs à l'instant $n-L$, avec $L = \max(p, q)$. Par conséquent, pour tous les chemins finissant, à l'instant n , par le même dernier sous-chemin $C(n-L+1\dots n)$, cette probabilité est la même. On ne garde donc que celui de probabilité $\Pr(C(0\dots n))$ maximale. Pour un quantificateur en loi A (256 niveaux de quantification), cet algorithme nécessite donc, dans une première approche, de mémoriser et actualiser à chaque échantillon 256^L chemins, avec les probabilités et bruits correspondants. La complexité est réduite de la manière suivante. D'une part, nous approchons le masque par un modèle ARMA d'ordres faibles, $p = 5$ et $q = 4$, ce qui assure un bon compromis entre la précision de la modélisation et la complexité. D'autre part, en considérant pour $w(n)$ une distribution gaussienne, la forte décroissance de la densité de probabilité autour de la moyenne permet de ne considérer, pour un chemin $C(0\dots n-1)$, que les K terminaisons les plus probables $Q(n)$ autour du centre de la gaussienne, au lieu des 256 possibles. Nous choisissons ici $K = 4$.

3 Résultats objectifs

Les deux méthodes sont simulées avec un modèle ARMA d'ordres 5 et 4 actualisé toutes les 16 ms pour approcher le seuil de masquage calculé selon la méthode de Johnston [3]. Elles permettent d'obtenir un bruit de quantification dont la forme du spectre suit bien le modèle fixé. En revanche, le niveau du bruit n'est pleinement contrôlé dans aucune des deux méthodes et dépasse occasionnellement le seuil de masquage, ce qui se traduit par un bruit rauque sporadique, plus ou moins fréquent selon les locuteurs et les sons prononcés.

Une première mesure des performances de masquage est donnée par la valeur λ effective (équation (1)), représentée pour deux locuteurs et pour les deux méthodes sur la figure 4. Un dépassement du modèle du masque se traduit par une valeur λ supérieure à 0 dB. Pour les valeurs de λ les plus critiques (supérieures ou légèrement inférieures à 0 dB), les performances de la méthode probabiliste sont meilleures, ce qui peut s'expliquer par le fait que l'optimisation du masquage n'y est pas contrainte.

Ce critère mesure cependant plus la capacité des algorithmes à reformer le bruit selon un modèle donné que le masquage proprement dit, le modèle utilisé n'étant qu'une approximation du masque. Or la complexité de la méthode probabiliste limite en pratique la finesse de la modélisation. Cette limite n'existe pas pour la première méthode, qui est alors simulée avec un modèle MA d'ordre 20, approchant plus précisément le masque. Nous comparons les deux méthodes selon un second critère plus représentatif du masquage, le rapport bruit à masque (RBM), défini par :

$$RBM = \frac{\sum_{i \in A} \gamma_b(i)}{\sum_{i=1}^{18} \gamma_{\text{Masque}}(i)}, A = \{i | \gamma_b(i) > \gamma_{\text{Masque}}(i)\} \quad (10)$$

où i désigne la i^{e} bande critique et $\gamma_b(i)$ la densité spectrale moyenne de bruit sur la i^{e} bande critique. Ce critère permet par ailleurs de comparer le reformage du bruit au cas sans reformage. Le rapport bruit à masque, qui n'est défini que pour les trames où le bruit dépasse le masque dans au moins une bande

critique, est représenté sur la figure 4 pour les deux mêmes locuteurs. Les parties visibles sur les courbes correspondent aux zones du signal où le bruit est susceptible d'être audible. Selon ce critère,

- aucune des deux méthodes ne se révèle meilleure que l'autre dans les simulations réalisées ;
- pour les deux méthodes, la fréquence d'audibilité du bruit reformé dépend des locuteurs, mais reste très inférieure à celle du bruit non reformé, avec un RBM inférieur dans les zones d'audibilité.

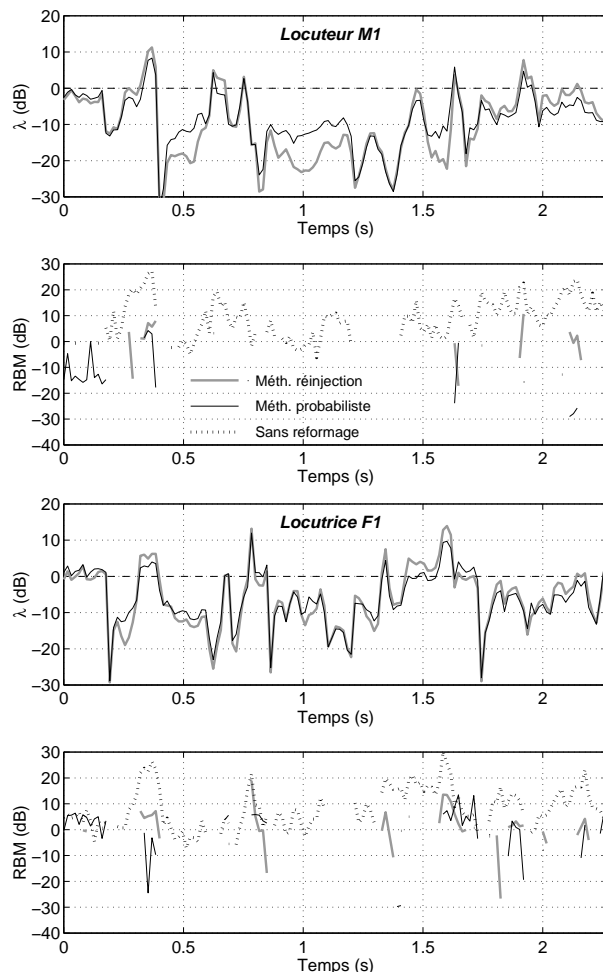


FIG. 4 – Écart λ entre le spectre du bruit reformé et le modèle AR,MA4 du masque ; rapport bruit à masque (RBM)

4 Évaluation subjective

L'évaluation subjective vise en premier lieu à comparer la perception du bruit de quantification reformé à celle du bruit non reformé. Ce bruit étant lié à l'égaliseur, nous évaluons d'autre part la préférence des auditeurs entre les signaux en réception d'une liaison égalisée, bruités, et le signal en réception de la même liaison non égalisée, non bruité.

Pour la majorité des locuteurs, ni les résultats objectifs ni les écoutes informelles réalisées sur les signaux de réception n'indiquent de supériorité d'une des deux méthodes de reformage en termes de masquage du bruit. C'est pourquoi, pour simpli-

fier les conditions de test, seule la méthode de réinjection de l'erreur de quantification a été évaluée formellement. Ce choix se justifie par ailleurs par ce que la méthode probabiliste induit, pour certaines locutrices, un léger sifflement dans le signal de réception.

Le niveau du bruit de quantification dépendant principalement du locuteur et des phonèmes prononcés, une seule liaison est simulée, pour :

- 4 locuteurs (2 hommes : M1 et M2 ; 2 femmes : F1 et F2) ;
- 2 doubles-phrases, d'une durée de 8 s chacune, pour chaque locuteur.

Pour chaque locuteur (locuteur, phrase), nous notons respectivement A, B et C les signaux de réception :

- de la liaison non égalisée ;
- de la liaison égalisée, sans reformage du bruit ;
- de la liaison égalisée, avec reformage du bruit.

Le test subjectif, effectué par 24 sujets, consiste en une comparaison par paires [2], dans laquelle, pour chaque paire de {A,B,C}, les auditeurs indiquent l'élément préféré (choix binaire). Chaque paire est présentée dans les deux sens, de manière à annuler l'effet de l'ordre de présentation.

Pour chaque locuteur et chaque paire {X,Y} de {A,B,C} nous calculons le pourcentage de préférence de X à Y. De ces pourcentages nous déduisons les positions relatives de A, B et C sur une échelle de préférences de Thurstone, tel que représenté sur la figure 5.

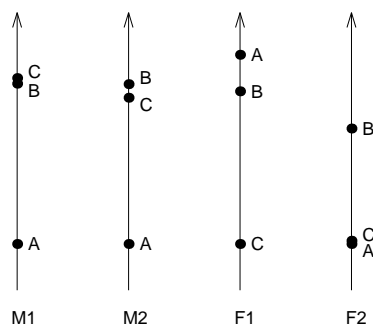


FIG. 5 – Echelles de Thurstone des préférences

Pour les locuteurs masculins, les auditeurs n'expriment pas de préférence entre les bruits reformés et non reformés, et préfèrent nettement (à 80 %) B ou C à A. Au contraire, le signal C (signal égalisé, avec reformage du bruit) est le moins apprécié dans le cas des locutrices, pour lesquelles B lui est préféré à 78 % (F1) et 72 % (F2).

Ainsi, bien que très sporadiquement non masqué, le bruit reformé selon les méthodes proposées est perçu par les auditeurs comme au mieux aussi gênant qu'un bruit de quantification blanc audible en permanence. Notons toutefois que, la problématique du bruit de quantification étant liée à celle de l'égalisation, il convient d'évaluer conjointement la perception du timbre de la voix et du bruit : à cet égard, les résultats pour les locuteurs M1,M2 et F2 ne disqualifient pas nos méthodes de reformage.

5 Conclusion

Nous avons proposé deux méthodes originales permettant de reformer spectralement le bruit d'un quantificateur de type MIC simplement par une modification de l'opération de quantification. Le bruit ainsi reformé dépasse peu fréquemment le seuil de masquage mais, subjectivement, ce bruit rauque sporadique est perçu comme au moins aussi gênant que le bruit non reformé. Ce bruit résultant de la prédistorsion du signal (égalisation dans le cas étudié), il pourrait y être remédié par une optimisation subjective conjointe de la pré-distorsion et du reformage du bruit.

6 Remerciements

Nous remercions chaleureusement Alain Le Guyader pour les enrichissantes discussions sur le masquage du bruit, ainsi que Lætitia Gros pour son aide précieuse dans la conception des tests subjectifs.

Références

- [1] R. Boite, H. Bourlard, T. Dutoit, J. Hancq et H. Leich , *Traitement de la parole*, Presses polytechniques et universitaires romandes, 2000.
- [2] C. Bonnet, *Manuel pratique de psychophysique*, Armand Colin, 1986, pp. 136-142.
- [3] J.D. Johnston, *Transform coding of audio signals using perceptual noise criteria*, IEEE Journal on selected areas in communications, vol. 6, n°2, pp. 314-323, février 1988.
- [4] G. Mahé et A. Gilloire, *Correction of the voice timbre distortions on telephone network*, Proc. Eurospeech, pp 1867-1870, septembre 2001.
- [5] G. Mahé et A. Gilloire, *Quantization noise spectral shaping in instantaneous coding of spectrally unbalanced speech signals*, Proc. IEEE Workshop on Speech Coding, Tsukuba, pp 56-58, octobre 2002.
- [6] J. Makhoul et M. Berouti, *Adaptive noise spectral shaping and entropy coding in predictive coding of speech*, IEEE Transactions on acoustics, speech, and signal processing, vol. ASSP-27, n°1, pp. 63-73, février 1979.
- [7] UIT-T, Recommandation P.830, *Évaluation subjective de la qualité des codecs numériques à bande téléphonique et à bande large*, Annexe D, février 1996.