Paramétrisation de la Parole basée sur une Modélisation des Filtres Cochléaires: Application au RAP

Zied Hajaiej, Kais Ouni, Noureddine Ellouze

Laboratoire des Systèmes et Traitement du Signal (LSTS)
Ecole Nationale d'Ingénieurs de Tunis, BP 37, Le Belvédère, 1002 Tunis, Tunisie
Hajaiej_zied@yahoo.fr
(Kais.ouni, N. Ellouze@enit.rnu.tn)

ABSTRACT

Signal processing front end for extracting the feature set is an important stage in any speech recognition system. The optimum feature set is still not yet decided. There are many types of features, which are derived differently and have good impact on the recognition rate. This paper presents one more successful technique to extract the feature set from a speech signal, which can be used in speech recognition systems. Our technique based on the human auditory system characteristics and relies on the gammachirp filterbank to emulate asymmetric frequency response and level dependent frequency response. For evaluation a comparative study was operated with standard MFCC and PLP.

1. INTRODUCTION

Dans les deux dernières décennies, les modèles numériques appliqués au système auditif périphérique ont gagné une popularité croissante en traitement des signaux de parole, en particulier en reconnaissance de la parole. Par ailleurs, les filtres roex " rounded exponential " présentent une bonne approximation des données expérimentales auditives [4], sous les hypothèses simplificatrices que les filtres auditifs sont symétriques sur une échelle logarithmique, et que leur étalement loin de la fréquence centrale fc est négligé. Néanmoins, ils sont définit dans le domaine spectral ce qui rend difficile leur implémentation selon le schéma conventionnel des structures de bancs de filtres auditifs. Pour palier cet inconvénient, un modèle temporel a été proposé pour la première fois par Johannesma [1], appelée gammatone. Ce modèle a l'avantage d'être définie par une réponse impulsionnelle temporelle, c'est un filtre à bande critique obéissent à la mesure de Bark et se présente sous la forme d'une enveloppe de type gamma modulée. Le filtre gammatone présente une enveloppe spectrale symetrique, or les données psychoacoustiques optent pour une enveloppe non symétrique où le degré d'asymétrie dépend du niveau sonore [5], Irino et Patterson ont proposé un nouveau modèle du filtre auditif qui dérive de la fonction gammatone, appelé gammachirp, pour introduire une dépendance vis à vis du niveau d'intensité du stimulus sonore appliqué.

Cette dépendance se présente sous la forme d'un paramètre supplémentaire dans l'expression de la gammatone qui génère l'asymétrie du spectre d'amplitude. Dans ce papier nous proposons en premier lieu deux techniques de paramétrisation des signaux de parole basées sur un banc de filtres gammachirp qui imite le comportement spectral de la cochlée, en suivant la démarche utilisée dans les techniques MFCC et PLP. En second lieu, l'approche adoptée pour l'étude de la validité des deux techniques proposées ainsi que leur évaluation par rapport aux techniques de paramétrisation standards MFCC et PLP. Le système de reconnaissance adopté pour cette étude est celui de HTK basé sur les modèles de Markov cachés HMM.

2. LES TECHNIQUES DE PARAMETRISATION

Il existe dans la littérature une grande variété de technique de parametrisation des signaux de la parole, nous citons les plus important qui on révolutionne en quelque sorte le domaine de la reconnaissance de parole a savoir MFCC et PLP.

2.1. Coefficients mel-cepstre (MFCC)

Cette technique consiste à calculer les coefficients cepstraux sur une échelle en Mel qui se rapproche de la perception fréquentielle de l'oreille. Après l'application d'une transformée de Fourier à court terme, l'énergie est calculée dans des brandes critiques modélisées par des filtres triangulaires quant à l'échelle des amplitudes est exprimée en décibels. L'échelle des fréquences quant à lui est exprimée en Mel. Le cepstre est ensuite calculé par l'expression suivante :

$$C_n = \sqrt{\frac{2}{k}} \sum_{k=1}^{N} \left(\log S_k \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{k} \right] \right) \qquad (1)$$

Avec k=1,…,N et S_k représentant l'énergie correspondante après filtrage par un $k^{ième}$ filtre triangulaire.

2.2.L'analyse prédictive linéaire perceptuelle (PLP)

La paramétrisation par prédiction linéaire (LPC) à pour principal défaut le fait d'estimer uniformément le spectre sur toutes les fréquences de la bande audible. Ainsi il est possible que certains détails spectraux ne soient pas pris en compte par la technique LPC ou encore qu'ils prennent une importance majeure sans qu'ils soient physiologiquement pris en compte par l'oreille. En effet, la technique PLP [9,11] permet de résoudre ce problème : l'analyse opérée par cette technique a pour but d'estimer des paramètres d'un filtre autorégressif tout pôle, modélisant au mieux le spectre auditif.

3. FILTRE GAMMACHIRP

Le filtre gammachirp est utilisé dans la recherche psychoacoustique comme étant un modèle fiable du filtre cochléaire. Il est défini dans le domaine temporel par la partie réelle de la fonction $g_c(t)$ [1, 3, 5].

$$g_c(t) = at^{n-1}\exp(-2\pi bERB(f_r)t)\times\exp(j2\pi f_r t + jc\ln t + jc\varphi) \quad (2)$$

Avec t>0, a paramètre de normalisation d'amplitude, f_r la fréquence de modulation, n l'ordre du filtre, $bERB(f_r)$ un paramètre définissant l'enveloppe du filtre. L'ERB représente quant à lui la largeur de bande rectangulaire équivalente [3,6].

$$ERB(f_r) = 24.7 + 0.108\, f_r \quad (3)$$

c représente un facteur introduisant l'asymétrie de ce filtre et φ la phase initiale, ln t est un logarithme népérien de temps, La figure 1 donne un exemple de réponse impulsionnelle du filtre gammachirp.

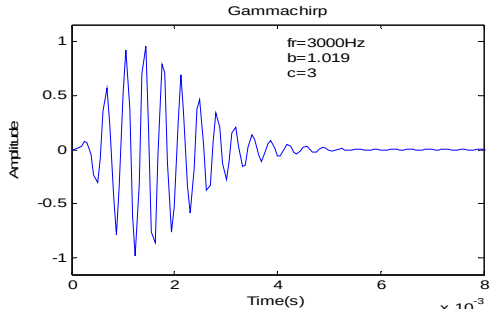

Figure 1 : Exemple de réponse impulsionnelle du filtre gammachirp.

La transformée de Fourier de la réponse impulsionnelle est donnée par l'équation suivante [3] :

$$G_c(f) = \frac{a|\Gamma(n+jc)|}{\Gamma(n)} \cdot \frac{\Gamma(n)}{|2\pi\sqrt{(b.ERB(f_r))^2 + (f-f_r)^2}|^n} e^{c\theta} \quad (4)$$

$$|G_c(f)| = a_\Gamma |G_T(f)| . e^{c\theta(f)} \quad (5)$$

$$\theta(f) = \arctan\left(\frac{f - f_r}{bERB(f_r)}\right) \quad (6)$$

Le transfert de gammachirp se présente comme le produit du transfert de la gammatone $G_T(f)$ par une fonction de transfert appelée fonction d'asymétrie $e^{c\theta(f)}$. Le degré d'asymétrie dépend de c, si c est négatif la fonction de transfert $e^{c\theta(f)}$ se comporte comme un filtre passe-bas et dans le cas ou C est positif elle se comporte comme un filtre passe-haut. Les études psychoacoustiques montre que c est fortement dépendant de la puissance du signal. En effet le paramètre c est relié à la puissance du signal par une expression de type $c=3.38-0.107P_s$ [3], avec P_s puissance de signal d'entré. Le pic de fréquence f_p de ce spectre est défini pour $G'(f_p)=0$. Ce pic est décalé de f_r par [4, 5].

$$f_p = \frac{f_r + cbERB(f_r)}{n} \quad (7)$$

Ce décalage est dû au paramètre c introduit dans l'expression de la réponse impulsionnelle gammachirp g_c. Ce décalage ainsi que l'asymétrie du spectre de la gammachirp représente une approximation intéressante aux résultats psychoacoustiques disponibles [2,7].

4. IMPLÉMENTATION DE BANC DE FILTRE GAMMACHIRP

La paramétrisation des signaux s'opère par un banc de filtre calculé à l'aide de la fonction gammachirp. Dans notre application, on utilise un banc de 32 filtres caractérisés par 32 réponses impulsionnelles gammachirp, où la fréquence central de chaque filtre gammachirp a une largeur de bande ERB et le banc couvre la bande 50-8000 Hz, Le signal de parole est segmente par une fenêtre de Hamming de largeur 25 ms. Chaque section du filtre gammachirp se compose de deux chemins, le premier chemin est le filtrage et le deuxième est l'estimation de la puissance du signal dans chaque sous bande. Le chemin de filtrage est un filtre gammatone d'ordre 4 suivi de la fonction d'asymétrie pour réaliser le filtre gammachirp final. Dans le deuxième chemin on calcule la puissance de signal dans chaque sous bande et c en utilisant l'expression suivante :

$c=3.38+0.107P_s$. La sortie du banc de filtre gammachirp est soumit à une opération de filtrage par la courbe d'égale intensité. La figure 2 donne la démarche utilisée pour déterminer le banc de filtre gammachirp, La figure 3 montre la réponse impulsionnelle de 32 filtres gammachirp, couvrant la bande de 50-8000 Hertz, après filtrage par la courbe d'égale intensité. La figue 4 donne les caractéristiques du banc de filtre gammachirp.

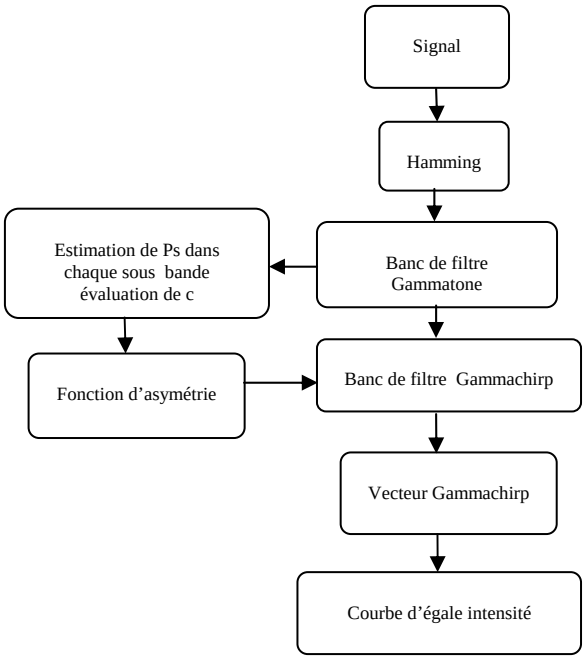

Figure 2: Bloc diagramme du filtre gammachirp.

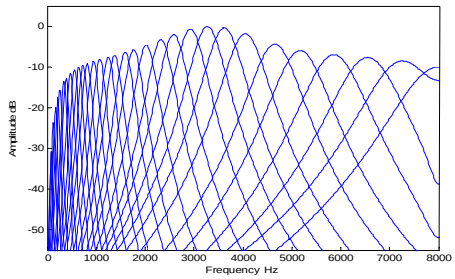

Figure 3: Exemple de banc de filtre gammachirp après filtrage par courbe d'égale intensité.

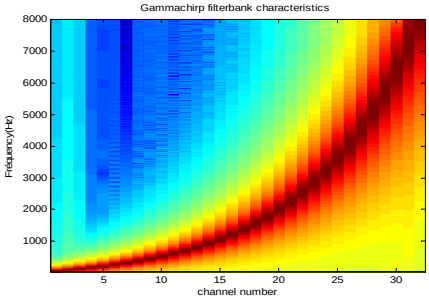

Figure 4: Caractéristique du banc de filtre gammachirp.

5. ANALYSE DE LA PAROLE BASEE SUR UN BANC DE FILTRE GAMMACHIRP

Les coefficients physiologiques du banc de filtre gammachirp peuvent être utilisé autant que les coefficients de paramétrisation du signal de parole, dans ce cas l'énergie dans chaque filtre est calculée par estimation du module de la transformée de Fourier discrète (DFT) du signal en la multipliant par le filtre gammachirp correspondant. De cette étape nous avons expérimenté deux options : La première option est GammaChirp Cepstral (GC-Cept) qui consiste à estimer les coefficients cepstraux par la transformée de cosinus discrète (DCT), ce qui décorele le signal de parole, en réduisant le nombre de coefficients d'analyse à 12 coefficients cela étant nécessaire pour le traitement du HMM. La deuxième option est GammaChirp-PLP (GC-PLP) qui est réalisée de sorte à suivre les étapes du PLP. La figure 5 donne Les différentes démarches de GC-Cept et GC-PLP.

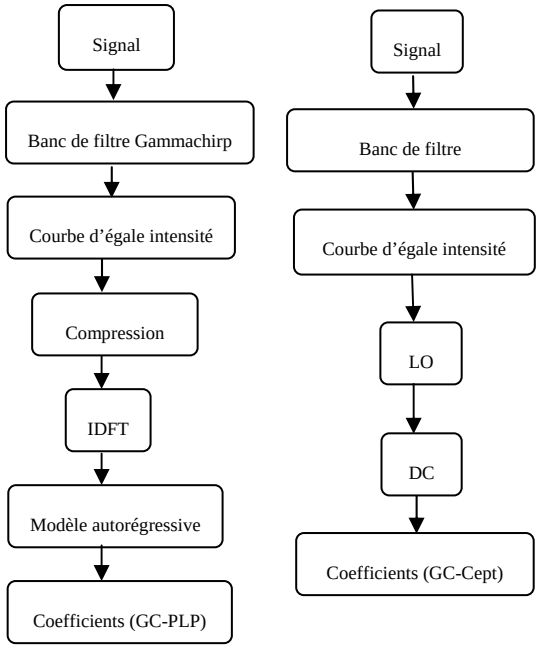

Figure 4: Parametrisation basée sur un banc de filtre gammachirp

6. EVALUATION

Pour évaluer le banc de filtre auditif gammachirp, nous avons comparé différentes techniques de paramétrisation standards (MFCC et PLP) avec les paramétrisations proposées (GC-Cept et GC-PLP) dans le cadre de la reconnaissance de mots isolés. Les 12 premiers coefficients auxquels on a ajouté l'énergie (soit 13 coefficients), les delta et les delta-delta (soit 39 coefficients). L'évaluation porte sur un corpus issu de la base TIMIT composé de 6132 mots pour la phase d'apprentissage, soit 21 mots répètes 292 fois par 18 hommes et 18 femmes réparties uniformément sur 8 dialectes américains. Pour la phase de reconnaissance nous avons utilisés 2201 mots, soit 21 mots répétés 104 fois par 13 hommes et 13 femmes réparties uniformément sur 8 dialectes américains. En ce qui concerne le modèle de Markov Cachés, HMM [9], nous avons utilisés une matrice de taille 5x5 pour l'ensemble des probabilités de transition.

Les probabilités d'émission sont de type multi gaussienne, définies uniquement par les vecteurs moyens de dimension 12, les matrices de covariance et

les poids associés à chaque gaussienne de dimension 12. Les résultats de chaque type de paramètrisation sont représentés, tout d'abord, dans un état brut, ensuite avec l'énergie, delta et delta-delta. Les tableaux 1, 2, 3 et 4 donnent les résultats associés aux taux de reconnaissances des différentes techniques de paramètrisation.
Nous définissions les paramètres si dessous.

N: Le nombre total de mots à reconnaître.
D: Le nombre de mots non pris.
S: Le nombre de mots non reconnus.
H: Le nombre de mots reconnus.
%: Le taux obtenu en pourcentage,

Table 1: Taux de reconnaissance obtenu par les techniques de paramétrisation dans leur état brut.

	%	N	H	S	D
MFCC	91.00	2201	2003	198	0
PLP	91.78	2201	2020	181	0
GC-PLP	94. 86	2201	2175	124	0
GC-Cept	91.86	2001	2025	186	0

Table 2 : Taux de reconnaissance obtenu par les techniques de paramétrisation combinées avec l'énergie.

	%	N	H	S	D
MFCC_e	93.14	2201	2050	151	0
PLP_e	94.14	2201	2072	129	0
GC-PLP_e	95.72	2201	2126	75	0
GC-Cept_e	95.20	2201	2112	89	0

Table 3 : Taux de reconnaissance obtenu par les techniques de paramétrisation combinées avec l'énergie et delta.

	%	N	H	S	D
MFCC_e_d	98.36	2201	2165	36	0
PLP_e_d	98.41	2201	2166	35	0
GC-PLP_e_d	98.94	2201	2184	17	0
GC-Cept_e_d	98.53	2201	2174	27	0

Table 4 : Taux de reconnaissance obtenu par les techniques de paramètrisation combinées avec l'énergie, delta et delta-delta.

	%	N	H	S	D
MFCC_e_d_a	98.64	2201	2171	30	0
PLP_e_d_a	99.05	2201	2180	21	0
GC-PLP_e_d_a	99.56	2201	2193	8	0
GC-Cept _e_d_a	99.10	2201	2182	19	0

7. CONCLUSION

Dans ce papier, nous avons présenté deux techniques de paramétrisations du signal de parole qui tient compte des caractéristiques fréquentielles de l'oreille, basée sur un banc de filtres dont les réponses impulsionnelles sont celles des fonctions gammachirp. Les paramétrisations implémentées ont montré leurs performances avec le système de reconnaissance automatique de la parole HTK basé sur le Modèles de Markov Cachés au vu d'une reconnaissance de mots isolés. Nous observons que les résultats les moins bons sont ceux obtenus avec la modélisation de base et les meilleurs sont ceux obtenus avec la modélisation avec banc de filtre gammachirp.

BIBLIOGRAPHIE

[1] K. Ouni. Contribution à l'analyse du signal vocal en utilisant des connaissances sur la perception auditive et représentation temps fréquence en multirésolution des signaux de parole .Thèse de Doctorat*, ENIT, 2003*.

[2] T. Irino, R. D. Patterson. Temporal asymmetry in the auditory system. *J. Acoust. Soc. Am*. 99(4): 2316-2331, April, 1997.

[3] T. Irino, D. Patterson. A time-domain, level-dependent auditory filter: the gammachirp. *J. Acoust Soc. Am.* 101(1): 412-419, January, 1997.

[4] T. Irino et M. Unoki. An analysis auditory filterbank based on an IIR implementation of the gammachirp. *J. Acoust. Soc Japan.* 20(6): 397-406, November, 1999.

[5] T. Irino, R. D. Patterson. A compressive gammachirp auditory filter for both physiological and psychophysical data. *J. Acoust Soc. Am.* 109(5): 2008-2022, may 2001.

[6] J. O. Smith III, J.S. Abel. Bark and ERB bilinear transforms, *IEEE Tran. On speech and Audio Processing*, Vol. 7, No. 6, November 1999.

[7] R. D. Patterson, I. Nimmo-Smith. Off-frequency listening and auditory-filter asymmetry, *J. Acoust. Soc. Am.,* Vol. 67, No. 1, pp. 229-245, 1980.

[8] Irino, T. and Unoki, M. (1998). A time-varying, analysis/synthesis auditory filterbank using the gammachirp. *IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-98)*, 3653-3656.

[9] H. Hermansky. Perceptual lineair predictive (PLP) analysis of speech. *J. Acoust. Soc. Am.* Vol. 87, No. 4,pp. 1738-1752., April 1990.

[10] B.R. Glasberg, B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research,* 47103-198, 1990.

[11] H. Hermansky, J. C. Junqua, Optimization of Perceptually Based ASR Front-Ends. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-88), New York*, April 11-14, 1988, paper S 5.10,pp.219-222.