Sur l'évaluation du second formant F'2 par une technique d'estimation spectrale basée sur une modélisation du filtrage auditif

Kaïs Ouni et Noureddine Ellouze

Laboratoire des Systèmes et Traitement du Signal (LSTS)
Ecole Nationale d'Ingénieurs de Tunis (ENIT)
BP.37, Le Belvédère, 1002, Tunis, Tunisie
Mél: kais.ouni@enit.rnu.tn - N.ellouze@enit.rnu.tn

ABSTRACT

In this paper, we propose a spectral estimation technique based on a gammachirp filterbank which is designed to provide a spectrum reflecting the spectral properties of the cochlea. The characteristic shift of the spectral peak of the gammachirp is then used to estimate perceptual formant F'2 of 18 cardinal vowels used by Bladon and Fant. We compare then the standard deviation of these results with those obtained by three traditional techniques. The first one suggested by Bladon and Fant, the second one by Paliwal et al., and the third one by Hermansky. The results show that the gammachirp spectral estimation gives a better estimate of F'2 than the second and the third techniques. It is a little less accurate than the first one.

1. INTRODUCTION

La plupart des études en traitement de la parole et en phonétique[Ste89][Del52][Nak95]ont montré que presque toutes les voyelles peuvent être simulées perceptiblement par des modèles à deux-formants, dans lesquels, l'un ou l'autre des deux formants constitue une moyenne pondérée des fréquences de deux ou plusieurs formants[Pol69]. Dans de telles études, des auditeurs ont été invités à apparier des configurations de deux-formants aux voyelles synthétiques standards formées de quatre formants ayant des fréquences appariant ceux des voyelles naturelles. Le premier formant de ce type de configuration à deux formants était fixé à la fréquence du F1 standard, et les sujets ont ajusté le deuxième formant F2' jusqu'à ce qu'ils aient entendu une voyelle qui est perçue comme équivalente à la voyelle originale composée de quatre formants. Dans le cas des voyelles postérieures, le F2' a été ajusté sur une intermédiaire de fréquence entre le F2 et le F3 de la voyelle originale. Dans le cas des voyelles antérieures, le F2' a été ajusté sur une fréquence près du F2. Pour expliquer ces observations, Delattre [Del52] suggère que quand les formants sont étroitement proche en fréquence comme le cas de F1 et F2 pour les voyelles antérieures et le F2 et F3 pour les voyelles postérieures, elles sont intégrés d'une façon perceptive tels que le formant pertinent F'2 est équivalent à une moyenne des formants standards. Chistovich [Chi79], a prouvée que quand deux crêtes spectrales ou plus se produisent dans une même bande critique dans l'échelle de Bark la qualité perçue de la voyelle est équivalente à une configuration avec une crête spectrale unique située au centre de gravité des fréquences des formants. La fréquence perceptive du formant F'2 est une moyenne pondérée en fréquence et en amplitude des crêtes spectrales dans une marge de 3-3.5 Barks. Dans cette marge, la fréquence perceptive du formant F'2 est décalée vers la fréquence de la crête la plus élevée en amplitude. Quand la distance de fréquence entre les crêtes spectrales excèdent 3.5 Barks, les formants sont perceptiblement éloignés et les changements de leur amplitude relative n'affectent pas la qualité perçue des voyelles.

Ces recherches nous ont amené à faire un rapprochement avec les propriétés de masquage spectral qui se manifeste dans la cochlée. Ce traitement spectral peut être simulé par un banc de filtres auditifs [Fla72]. Des filtres avec une réponse impulsionnelle de type gammatone sont largement utilisés pour modéliser le banc de filtres cochléaire. Il s'agit d'une fonction qui a l'allure des distributions gamma modulé par un sinus [Nak95]. Elle est distinguée par une largeur de bande spectrale qui dépend de la fréquence centrale de son filtre cochléaire correspondant, qui est mesurée en largeur de bande rectangulaire équivalente (ERB) [Oun01]. L'ERB est lié à la notion de bande critique issue d'expériences psychoacoustiques. La largeur de bande critique est d'environ 40-100 Hertz en basses fréquences et change graduellement en environ 20 pour cent de la fréquence en hautes fréquences. Récemment, Irino [Iri97] a proposé un nouveau modèle appelé Gammachirp pour le filtre cochléaire tenant compte de son asymétrie naturelle et qui dépend de l'intensité du signal. C'est une extension du filtre gammatone avec l'introduction d'un terme de décalage pour produire l'asymétrie du spectre d'amplitude.

Dans ce travail, un banc de filtres gammachirp est conçu pour fournir un spectre reflétant les propriétés spectrales de la cochlée, avec une approche semblable à l'estimation lissée et modifiée de spectre [Oun01]. Le décalage caractéristique du pic spectral de la gammachirp est ensuite utilisé pour estimer F'2. Une comparaison des résultats obtenus par la technique proposée et ceux collectés par Bladon et Fant [Bla78][Bla83] ainsi que les techniques classiques d'estimation de F'2 est enfin donnée et commentée. Dans les paragraphes suivants, nous présentons la technique proposée pour l'estimation spectrale par gammachirp. Puis, nous présentons

l'approche de l'estimation F'2 et nous terminons par l'évaluation de cette technique et une conclusion.

2. TECHNIQUE PROPOSÉE D'ESTIMATION SPECTRALE PAR GAMMACHIRP

2.1 Le filtre gammachirp

Le filtre gammachirp est une bonne approximation du comportement spectral et sélectif de la cochlée, il est définit dans le domaine temporel par la partie réelle de la fonction complexe $g_c(t)$ [Iri97].

$$g_c(t) = A\, t^{n-1} \exp(-2\pi B t) \exp\left(j 2\pi f_0 t + j c \ln(t) + j\varphi\right) \quad (1)$$

pour $t > 0$ et avec $B = b.ERB(f_0)$ (2)

Le terme n représente l'ordre du filtre, f_0 la fréquence de modulation, A représente une constante de normalisation de l'amplitude, b est un paramètre définissant l'enveloppe du filtre, c représente un facteur introduisant l'asymétrie de ce filtre et φ est la phase initiale. L'ERB représente quant à lui la largeur de bande rectangulaire équivalente. Pour des niveaux d'intensités modérés, l'ERB peut s'exprimer par l'équation suivante [Iri 97] :

$ERB(f_0) = 24.7 + 0.108 f_0$ en Hz (3)

et le pic de fréquence f_p du spectre d'amplitude est décalé de f_0 par : $f_p = f_0 + \frac{c.B}{n}$ (4)

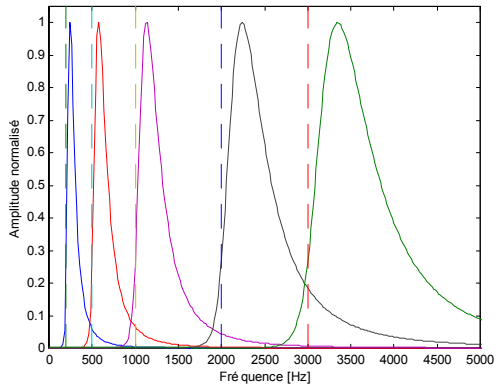

Figure 1: Exemple de spectre d'amplitude de gammachirp (*n=3*, *b=1* et *c=3*). Le décalage du pic spectral est proportionnel à la fréquence centrale de la gammachirp.

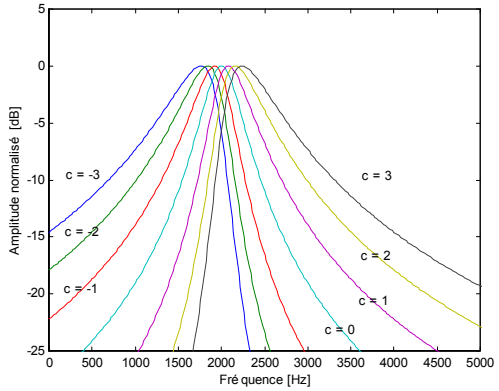

Figure 2: Exemples de spectres d'amplitudes normalisés et centrés sur la fréquence 2000 Hz en fonction de paramètre c . L'asymétrie de ces filtres est proportionnelle au paramètre c.

2.2 Estimation spectrale par gammachirp

Ce paragraphe présente une méthode pour l'estimation spectrale basée sur un banc de filtres gammachirp. Une approche similaire à l'estimation lissée et modifiée de spectre permet d'obtenir un spectre lissé et moyenné dans le temps. L'estimateur modifié obtenu est la moyenne temporelle de l'estimateur spectral dans chaque section. Il est donné par l'expression suivante [Oun01]:

$$R_{x_k}(f) = \left| \sum_{l=0}^{M-1} x_k(l) \,.\, g_c(l) \exp\left(-j 2\pi f l\right) \right|^2 \quad (5)$$

où $K = L/M$ est le nombre de sections du signal x, L est la longueur du signal et M est la longueur de chaque section. Le banc de filtres gammachirp ainsi formé est basé sur 512 filtres, répartis d'une façon linéaire sur l'échelle des fréquences. Un moyennage temporel est ensuite appliqué pour fournir un spectre unique représentant la contribution globale de tous les filtres.

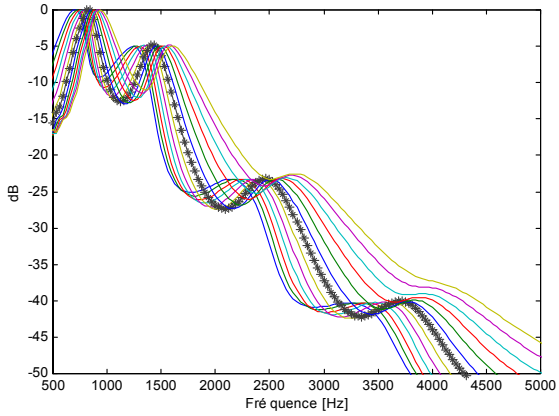

Figure 3: Estimations spectrales par gammachirp de la voyelle cardinale /a/ (Table1) pour un paramètre c variant dans l'ordre de gauche à droite de –3 à +3 avec un pas de 0,5, tout en fixant *b=1* et *n=3*. La courbe en étoile représente le spectre obtenu pour *c=0*.

3. APPLICATION DE LA TECHNIQUE PROPOSÉE POUR L'ESTIMATION DE F'2

3.1 Techniques d'estimations de F'2

Les méthodes d'estimation des formants perceptifs et essentiellement F'2 peuvent être classées dans deux catégories. La première calcule F'2 comme moyenne de l'ensemble des fréquences des quatre premiers formants. La deuxième estime F'2 sans utiliser les fréquences standards des formants. Ainsi, Bladon et Fant [Bla78][Bla83], Paliwal, Lindsay et Ainsworth [Pal83] et Carlson , Fant et Granstrom [Pal83] proposent des formules pour estimer le F2' comme moyenne pondérée des quatre premiers formants. Itahashi et Yokoyama

[Ita78] estiment F'2 par un spectre LPC d'ordre élevé modifié en échelle Mel et appliquant un seuillage d'égale sonie. Hermansky [Her90][Hu97] a utilisé une technique de prédiction linéaire perceptive (PLP) basé sur un modèle LPC de 5ème ordre dans un contexte auditif. L'estimation spectrale par gammachirp utilisée pour estimer les formants peut être incluse dans la deuxième catégorie puisqu'elle n'a pas besoin des fréquences des quatre premiers formants.
Dans un travail précédent [Oun01] nous avons utilisé la technique d'estimation spectrale par gammachirp pour estimer les formants réels en annulant le décalage provoqué par le paramètre *c*. Les résultats obtenus ont montré que cette technique estime correctement les formants. Dans le présent travail, nous posons l'hypothèse que le décalage observé dans la détermination de F'2 dans les expériences de Bladon et Fant [Bla78][Bla83] peut être expliqué par le paramètre *c* introduit dans la réponse impulsionnelle gammachirp.

3.2 Résultats et discussions

Pour valider l'approche proposée, nous avons utilisé les 18 voyelles cardinales analysées par Bladon et Fant [Bla78]. Dans leurs expériences, ils ont déterminé les valeurs de F'2 de 18 voyelles cardinales perçu par un nombre donné d'auditeurs. Nous proposons dans ce paragraphe de comparer les valeurs des fréquences obtenues par la technique d'estimation spectrale par gammachirp à ceux données dans [Bla78][Bla83].

Pour chaque voyelle, la technique d'estimation spectrale par gammachirp est appliquée pour obtenir le spectre correspondant. Un algorithme de détection de pics est appliqué ensuite pour fournir le deuxième pic spectral qui va être pris par hypothèse comme étant le F'2. En fixant le paramètre b = 1 et n = 3 et en faisant varier le paramètre c, l'écart type entre les valeurs données par Bladon et Fant et ceux estimées par la technique proposée ont donné un écart minimal pour $c=0.3$. De même nous avons comparé les résultats obtenus à trois techniques classiques proposés par Bladon et Fant (BF)[Bla78], Paliwal, Lindsay et Ainsworth (PLA) [Pal83] et par Hu et Barnard [Hu97] basé sur la technique de PLP mais avec un ordre de 6. Les résultats montrent que la technique proposée donne un meilleur écart type de l'ordre de 45,40 Hz comparés à ceux donnée par PLA et PLP et voisine celui donné par BF. Cet écart type est par coïncidence au même ordre de grandeur de la première bande critique. Dans le cas de la voyelle /i/, l'écart important observé pour le F'2 estimé par gammachirp par rapport à la valeur perceptive correspondante, peut s'expliquer par le fait que la technique proposée s'apparente toujours au second formant standard et au décalage du deuxième pic spectral de la gammachirp. Par contre pour la voyelle /ɯ/ l'estimation spectrale par gammachirp donne la meilleure estimation de F'2, même par rapport au modèle BF. La Table1 présente les fréquences des quatre premiers formants utilisées pour synthétiser les 18 voyelles cardinales, les valeurs de F'2 estimées par les trois techniques classiques citées dans le paragraphe précédent, et les valeurs de F'2 estimée par l'estimation spectrale par gammachirp. Un écart type est donné pour chaque technique. La figure 4 donne enfin la répartition des 18 voyelles cardinales dans le plan F1-F'2 mesurée en ERB, par les données collectées par Bladon et Fant comparée à ceux estimées par la technique d'estimation spectrale gammachirp et le plan classique F1-F2.

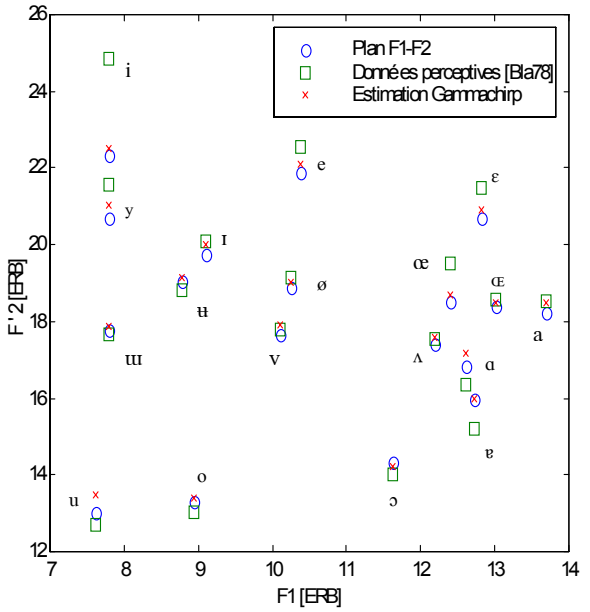

Figure 4 : Plan F1-F'2 des 18 voyelles cardinales estimées par l'estimation spectrale gammachirp comparé aux données perceptives [Bla78] ainsi que le plan F1-F2 classique.

4. CONCLUSION

Dans ce travail nous avons proposé une technique d'estimation spectrale basée sur un banc de filtres de type gammachirp qui présente une meilleure approximation des filtres cochléaires. Nous avons émis l'hypothèse que les valeurs décalées de F'2 perçues dans les expériences de Bladon et Fant peuvent être expliquées par le décalage caractéristique du pic spectral de la gammachirp. Pour valider cette hypothèse, nous avons procédé à l'estimation spectrale par gammachirp du deuxième formant des même voyelles cardinales utilisées par Bladon et Fant et nous avons calculé l'écart type entre les valeurs obtenues par la technique proposée et ceux issues de leurs expériences pour plusieurs valeurs du paramètre c qui commande le décalage du pic spectral de la gammachirp, tout en fixant l'ordre du filtre n à 3 et le paramètre b à 1. Les résultats ont donné un écart type optimal pour la valeur de c égale à 0,3. De même, nous avons comparé cette technique à trois techniques classiques proposées par Bladon et Fant, Paliwal, Lindsay et Ainsworth et par Hu et Barnard. Les résultats ont montré que la technique d'estimation spectrale par gammachirp donne une meilleure estimation de F'2 par rapport à la technique de Paliwal, Lindsay et Ainsworth et

à la technique de Hu et Barnard et voisine celle de Bladon et Fant.

Comme perspective de ce travail, nous nous proposons d'étudier l'ordre *n* et le paramètre *b* optimaux pour l'évaluation de F'2.

Table 1 : Table récapitulative des résultats obtenus.

Voyelle	Les quatre premiers Formants				Données Perceptives [Bla78]	Résultats Comparatifs			Valeurs estimées de F'2 par gammachirp
	F1	F2	F3	F4	F2'	F'2 (BF)	F'2(PLA)	F'2(PLP)	n=3; c=0,3
i	300	2300	3070	3590	3095	3190	2458	2215	2350
e	470	2180	2720	3790	2361	2361	2393	1963	2239
ɛ	680	1890	2580	3940	2076	1932	2107	1830	1944
a	770	1400	2460	3710	1452	1410	1679	1375	1447
ɑ	660	1170	2770	3650	1103	1182	1441	1255	1225
ɔ	570	840	2640	3310	806	842	1237	1210	829
o	370	730	2670	3240	700	733	945	1080	737
u	290	700	2550	3280	669	700	863	853	746
y	300	1890	2250	3000	2101	2125	2129	1963	1972
ø	460	1520	2290	3290	1570	1583	1691	1560	1539
œ	640	1450	2330	3030	1637	1612	1668	1505	1483
Œ	700	1430	2390	3350	1458	1471	1671	1452	1447
ɐ	670	1050	2900	3490	947	1072	1393	1233	1050
ʌ	620	1260	2390	3610	1284	1266	1490	1327	1290
ɤ	450	1300	2640	3470	1326	1354	1453	1426	1345
ɯ	300	1320	2480	3440	1300	1359	1440	1452	1336
ɪ	380	1690	2460	3570	1754	1763	1844	1797	1741
ʉ	360	1550	2430	3030	1503	1979	1692	1675	1566
Ecart Type [Hz]						29.89	61.85	69.13	45.40

BIBLIOGRAPHIE

[Bla78] Bladon A. and Fant G., "A two-formant model and the cardinal vowels," STL-QPRS, no 1, pp. 1-8, 1978

[Bla83] Bladon A. (1983), " Two-formant models of vowels perception: shortcomings and enhancements", Speech Communication, Vol. 2, No. 4, décembre, pp. 305-313.

[Chi79] Chistovich L. A. and Lublinskaja V. V., "The center of gravity effect in vowel spectra and critical distance between the formants : psycho-acoustical study of the perception of vowel-like stimuli", Hear. Res.,Vol.1, 1979, pp. 185-195.

[Del52] Delattre, P. C., Liberman, A. M., Cooper, F., and Gerstman, L. J., " An experimental study of the acoustical determinants of vowels colour", Word,8:195-210,1952.

[Fla72] Flanagan J.L. (1972), Speech Analysis, Synthesis and Perception, Springer-Verlag.

[Her90] Hermansky H., " Perceptual linear predictive (PLP) analysis of speech", J. Acoust. Soc. Am., Vol. 87, No. 4, April 1990, pp. 1738-1752.

[Hu97] Hu Z. and Barnard E., " Efficient estimation of perceptual features for speech recognition", Eurospeech 97, Greece, 1997, pp. 493-496.

[Iri97] Irino, T., Patterson, R. D., "A time-domain, level-dependent auditory filter : The gamma-chirp", JASA, Vol. 101, No. 1, pp. 412-419, January 1997.

[Iri01] Irino, T., Patterson R. D., "A compressive gamma-chirp auditory filter for both physiological and psychophysical data", J. Acoust. Soc. Am. Vol. 109, N° 5, Pt. 1, May 2001. pp. 2008-2022.

[Ita78] Itahashi S. and Yokoyama S., "A formant extraction method utilizing mel scale and equal loudness contour," STL-QPRS, no.4, pp.17-29, 1978.

[Nak95] Nakagawa, S., Shikano, K., and Tohkura, Y., " Speech, hearing and neural network models", *Edition Ohmsha IOS Press, 1995.*

[Oun01] Ouni k., Lachiri Z. and Ellouze N., "Formant Estimation using Gammachirp Filterbank", Eurospeech 2001 Scandinavia Conference, Denmark, September 3-7, 2001, pp. 2471-2474

[Pal83] Paliwal K. .K, Ainsworth W. A. and Lindsay D., "A Study of two-formant models for vowel identification", Speech Communication, Vol. 2, No. 4, décembre, pp. 305-313.

[Pic99] Pickett J. M., "The Acoustics of Speech Communication", Boston, MA: Allyn & Bacon, 1999.

[Pol69] Pols, L. C. W., LJ.T., van der Kamp and R.,

Plomp, “Perceptual and physical space of vowel sounds”, J. Acous. Soc. Am., Vol. 46, 1969, pp. 458-467.

[Ste89] Stevens K. N.,“On the quantal nature of speech”, Journal of Phonetics, Vol.17, 1989, pp. 3-45.