

Modélisation B-spline de contours mélodiques avec estimation du nombre de paramètres libres par un critère MDL

Damien Lolive, Nelly Barbot, Olivier Boëffard

IRISA / Université de Rennes 1 - ENSSAT
6 rue de Kerampont, B.P. 80518, F-22305 Lannion Cedex
damien.lolive,nelly.barbot,olivier.boeffard@irisa.fr
<http://www.irisa.fr/cordial/>

ABSTRACT

This article describes a new approach to estimate F_0 curves using a B-Spline model characterized by a knot sequence and associated control points. The free parameters of the model are the number of knots and their location. The free-knot placement, which is a NP-hard problem, is done using a global MLE within a simulated-annealing strategy. The optimal knots number estimation is provided by MDL methodology. Two criteria are proposed considering control points as real coefficients with variable precision. They differ on the precision used. Experiments are conducted in a speech processing context on a 7000 syllables french corpus. We show that a variable precision criterion gives good results in terms of RMS error (0.42Hz) as well as in terms of B-spline freedom number reduction (63% of the full model).

1. INTRODUCTION

Les technologies liées au traitement de la parole font largement appel aux modèles d'intonation. Notamment, la synthèse de la parole à partir du texte, TTS - Text-to-Speech systems -, ne peut se dispenser de tels modèles de manière à prédire des contours mélodiques à partir du texte et du style oratoire. Plus récemment, ces modèles ont été utilisés afin de déterminer une séquence optimale d'unités acoustiques prenant en compte des caractéristiques prosodiques[7]. Bien que l'intonation résulte d'une combinaison de nombreux facteurs linguistiques, cet article traite exclusivement du paramètre acoustique reconnu pour être le principal facteur de perception de la prosodie, à savoir la fréquence fondamentale ou F_0 . Les contours de F_0 , extraits du signal de parole, résultent de l'évolution au cours du temps de la vibration des cordes vocales. Une littérature importante traite de la modélisation de ces contours. On peut citer en particulier les modèles MoMel [4], Tilt [9], ainsi que les travaux de Sakai et Glass [8] qui utilisent des fonctions splines régulières.

Dans cet article, nous proposons une stylisation B-spline qui intègre une notion de régularité locale sans affecter l'approximation globale de la courbe F_0 , estimée selon un critère des moindres carrés. Pour conduire des expériences d'évaluation, nous avons dû faire le choix d'une unité prosodique minimale. Nous choisissons la syllabe comme support minimal d'un contour mélodique. Bien entendu, la méthodologie de stylisation proposée ne fait aucune hypothèse de cette nature. Dans [1], une comparaison entre les capacités de modélisation de F_0 par des splines et des B-splines est présentée selon un critère des moindres carrés. Les paramètres du modèle B-spline sont le nombre de nœuds et

leur placement. Un placement libre des nœuds permet de suivre les irrégularités des contours. Il est mis en œuvre par un algorithme de type Monte-Carlo (recuit simulé) afin de contourner la difficulté combinatoire de ce problème. Notre principal objectif est alors d'estimer le nombre optimal de nœuds afin de déterminer une classe de courbes B-splines parcimonieuse. Le cadre que nous avons choisi est celui du critère de longueur de description minimale, Minimum Description Length - MDL -, qui offre un compromis efficace entre la précision du modèle et le nombre de ses paramètres.

Dans la section 2, on introduit le modèle B-spline et ses paramètres estimés dans la section 3 au sens des moindres carrés. Dans la section 4, on présente les critères MDL pour optimiser le nombre de paramètres du modèle. Dans la section 5, le protocole expérimental est décrit et les résultats sont donnés dans la section 6.

2. MODÉLISATION B-SPLINE

Dans ce paragraphe, on présente les fonctions splines et leur généralisation aux courbes B-splines qui ont la capacité de modéliser des courbes ouvertes ayant des régularités locales variables.

Soit $[a, b]$ un intervalle que l'on subdivise en $l + 1$ sous-intervalles : $a = t_m < \dots < t_{m+l+1} = b$. Une spline de degré m associée à (t_i) est une fonction polynomiale de degré m sur chaque sous-intervalle et de classe $\mathcal{C}^{m-1}$ sur $[a, b]$. Les l points de jonction sont appelés des nœuds internes. L'ensemble de ces splines forme un espace vectoriel de dimension $m + l + 1$.

Dans [4], l'algorithme MoMel a pour but de fournir une stylisation de contours mélodiques à l'aide d'une spline quadratique. Pour cela, il estime n points cibles constituant les points stationnaires d'une spline de degré 2. Les points t_i correspondent aux abscisses des n points cibles et à leurs $(n - 1)$ points médians. Ainsi, la spline quadratique est associée à $l = 2n - 3$ nœuds internes et est entièrement déterminée par les $2n$ contraintes dues aux n points à interpoler avec une tangente horizontale.

L'espace vectoriel des splines de degré m associées à (t_i) admet une base de fonctions B-splines. Pour les définir, on pose $t_0 = \dots = t_m = a$ et $t_{m+l+1} = t_{2m+l+1} = b$, appelés nœuds externes, et on note $\mathbf{t}$ le vecteur contenant les nœuds internes et externes. Les B-splines sont alors définies par récurrence : au degré 0, pour i de 0 à $2m + l$

$B_0^i = \mathbf{1}_{[t_i, t_{i+1}[}$ et au degré m , pour i de 0 à $m + l$

$$B_m^i(t) = \frac{t - t_i}{t_{i+m} - t_i} B_{m-1}^i(t) + \frac{t_{i+m+1} - t}{t_{i+m+1} - t_{i+1}} B_{m-1}^{i+1}(t)$$

où les quotients sont nuls si $t_i = t_{i+m}$ ou $t_{i+1} = t_{i+m+1}$. Ajoutons que les B-splines sont positives et vérifient

$$\forall t \in [a, b[, \sum_{i=0}^{m+l} B_m^i(t) = 1. \quad (1)$$

Par conséquent, une fonction spline de degré m s'écrit comme une combinaison linéaire de fonctions B-splines de même degré. La généralisation des splines aux courbes B-splines autorise des nœuds internes non distincts et leur multiplicité désigne le nombre de fois où ils apparaissent dans $\mathbf{t}$. Ainsi, une courbe B-spline de degré m associée à un vecteur nœud $\mathbf{t}$ s'écrit comme une combinaison linéaire des fonctions B-splines de degré m dont les coefficients $c_0, \dots, c_{m+l}$ sont appelés points de contrôle.

On peut représenter matriciellement les points d'une courbe B-spline. Soit $x_0, \dots, x_{N-1}$ une suite de valeurs de $[a, b]$, on définit la matrice $\mathbf{B}$ qui a $B_m^i(x_j)$ comme (j, i) -ème élément, et on a, pour tout j de 0 à $N - 1$

$$g_{\theta^l}(x_j) = \sum_{i=0}^{m+l} c_i B_m^i(x_j) = (\mathbf{B}\mathbf{c})_j \quad (2)$$

où $\mathbf{c}$ désigne le vecteur colonne contenant les points de contrôle et $\theta^l = (t_{m+1}, \dots, t_{m+l}, c_0, \dots, c_{l+m})$.

Le principal effet des nœuds sur la courbe g_{θ^l} est lié à leur multiplicité. Soit t_i un nœud interne, plus sa multiplicité m_i est élevée et moins la courbe B-spline est régulière au point t_i . Plus précisément, si g_{θ^l} est de classe $\mathcal{C}^{m-1}$ entre deux nœuds consécutifs, elle est de classe $\mathcal{C}^{m-m_i}$ au nœud t_i . Quant aux points de contrôle, ils ont un impact local. En effet, si l'on fait varier c_i , cela modifie le terme $c_i B_m^i$ dans (2) et B_m^i étant nulle en dehors de $[t_i, t_{i+m+1}[$, seul le segment de courbe correspondant est affecté.

3. ESTIMATION DES PARAMÈTRES

Soit $\{(x_j, y_j), j = 0, \dots, N - 1\}$ un ensemble de mesures d'un contour mélodique, où (x_j) forme une suite croissante. On cherche à déterminer la courbe B-spline g_{θ^l} de degré m minimisant l'erreur quadratique moyenne avec les observations précédentes. Dans cette partie, on suppose le nombre l de nœuds internes connu, et on pose $t_0 = t_m = x_0$ et $t_{m+l+1} = t_{2m+l+1} = x_{N-1}$. On optimise dans un premier temps les points de contrôle, puis on estime la localisation des nœuds internes.

3.1. Les points de contrôle

Dans ce paragraphe, le vecteur nœud $\mathbf{t}$ est supposé connu et on détermine la courbe B-spline g_{θ^l} de degré m , i.e. ses points de contrôle, minimisant l'erreur quadratique avec les données. On note $\mathbf{y}$ le vecteur colonne de coordonnées y_j . D'après (2), on estime les points de contrôle tels que

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{y} - \mathbf{B}\mathbf{c}\|_2^2 = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y} \quad (3)$$

si $\mathbf{B}^T \mathbf{B}$ est inversible, i.e. lorsque les colonnes de $\mathbf{B}$ sont linéairement indépendantes. Pour cela, aucune multiplicité de nœud ne doit être supérieure à $m + 1$, sous peine d'avoir une colonne nulle. De plus, le nombre N de lignes de cette

matrice doit être nettement supérieur au nombre de colonnes afin de bien différencier chaque colonne de $\mathbf{B}$, i.e.

$$N \gg m + l + 1. \quad (4)$$

3.2. Placement des nœuds

On souhaite à présent déterminer un placement optimal des nœuds internes. Pour cela, on introduit le critère du maximum de vraisemblance et on choisit une stratégie de type Monte-Carlo (recuit simulé) pour déterminer une solution $\hat{\mathbf{t}}$ à ce problème NP-difficile. On considère que les nœuds se situent à des endroits d'observations, ils peuvent alors être représentés par un entier entre 0 et $N-1$.

On note e_j l'erreur commise entre l'observation y_j à l'instant x_j et sa modélisation $g_{\theta^l}(x_j)$. Afin de simplifier les calculs, on suppose $e_0, \dots, e_{N-1}$ indépendantes et distribuées selon une loi gaussienne centrée de variance σ^2 . Par conséquent, la log-vraisemblance du modèle s'écrit

$$\log p(\mathbf{y}; \theta^l) = -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{B}\mathbf{c}\|_2^2 - \frac{N}{2} \log(2\pi\sigma^2). \quad (5)$$

On remarque que $\hat{\mathbf{c}}$, défini par (3), est l'estimateur du maximum de vraisemblance pour $\mathbf{c}$. L'estimateur $\hat{\mathbf{t}}$ du maximum de vraisemblance pour $\mathbf{t}$ vérifie

$$\hat{\mathbf{t}} = \arg \min_{\mathbf{t}} \|\mathbf{y} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y}\|_2^2$$

où la matrice $\mathbf{B}$ dépend du paramètre $\mathbf{t}$.

Afin de déterminer $\hat{\mathbf{t}}$, nous avons mis en œuvre un algorithme d'optimisation globale selon une stratégie de type recuit-simulé (Simulated Annealing SA), présenté dans [1]. La relation proposée entre les paramètres du modèle et les distributions de l'algorithme SA est la suivante :

1. SA échantillonne un vecteur $\mathbf{v}$ de $\{x_1, \dots, x_{N-2}\}^l$.
2. Un vecteur $\mathbf{t}^*$ de nœuds internes est défini en triant les coordonnées de $\mathbf{v}$.
3. Un vecteur nœud $\mathbf{t}$ est défini en ajoutant $(m + 1)$ fois x_0 et x_{N-1} aux extrémités de $\mathbf{t}^*$.
4. Si deux nœuds consécutifs se trouvent dans un intervalle inférieur à 5% de $[x_0, x_{N-1}]$, ils sont fusionnés et la multiplicité du premier est incrémentée de 1.

La prochaine étape concerne l'optimisation du nombre l de nœuds internes. Avant de l'aborder, résumons les choix effectués quant au modèle B-spline. Pour l fixé, on commence par estimer un vecteur-nœud $\hat{\mathbf{t}}$ optimal à l'aide d'un algorithme SA. À partir de la matrice $\mathbf{B}$ associée et de (3), on calcule les points de contrôle $\hat{\mathbf{c}}$ optimaux. Par la suite, on désigne par $\hat{\theta}^l$ le modèle estimé.

4. B-SPLINES ET MDL

Le critère de longueur de description minimale (MDL) permet d'établir un compromis entre la qualité d'un modèle $\hat{\theta}^l$ et sa complexité l . Plus précisément, son application consiste à déterminer $\hat{l}$ minimisant la longueur de description des données $\mathbf{y}$. D'après [3],

$$\hat{l} = \arg \min_l L(\mathbf{y}) = \arg \min_l L(\hat{\theta}^l) - \log_2 p(\mathbf{y}; \hat{\theta}^l)$$

où $L(\hat{\theta}^l)$ et $-\log_2 p(\mathbf{y}; \hat{\theta}^l)$ désignent respectivement les longueurs de description du modèle estimé et des ob-

servations sachant ce dernier. L'expression (5) fait intervenir la variance σ^2 de l'erreur. Son estimateur selon le maximum de vraisemblance s'écrit :

$$\widehat{\sigma^2} = \arg \max_{\sigma^2} \log p(\mathbf{y}; \widehat{\theta}^l) = \frac{1}{N} \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_2^2.$$

Par conséquent, l'écart-type σ est estimé par la racine carrée de la moyenne des carrés des erreurs entre les observations et leur modélisation B-spline, également appelée erreur RMS (Root Mean Square). En injectant cette estimation dans (5), l'expression ci-dessus devient

$$L(\mathbf{y}) = L(\widehat{\theta}^l) + \frac{N}{2} + \frac{N}{2} \log_2(2\pi) + N \log_2(RMS).$$

4.1. Principe de solution

Considérons la longueur de description du vecteur $\widehat{\theta}^l$. Celui-ci est composé de l nœuds internes et $(l + m + 1)$ points de contrôle. On suppose que tous les nœuds et les points de contrôle sont respectivement de même longueur de description. Les nœuds sont positionnés dans $[x_0, x_{N-1}]$ à des endroits d'observations. Un nœud est alors représenté par un entier entre 0 et $N - 1$ et $\log_2(N)$ bits suffisent à son codage.

Quant aux points de contrôle, ce sont des paramètres réels estimés à partir de N données. Lorsque N est grand, la longueur d'un paramètre réel est généralement approchée par $\log_2(\sqrt{N})$. Cependant, chaque point de contrôle ayant une influence locale sur le modèle $g_{\widehat{\theta}^l}$, et le nombre d'observations pouvant être faible, cette approximation ne semble pas adaptée [2]. On considère alors une longueur de description d'un point de contrôle basée sur une loi a priori uniforme sur un intervalle borné $[-\alpha, \alpha]$ [5]. Pour une précision ε sur la description d'un point de contrôle, sa longueur est donnée par $\log_2(\alpha) + 1 - \log_2(\varepsilon)$. Pour $\widehat{\mathbf{t}}$ connu, l'estimation du maximum de vraisemblance pour α est $\|\widehat{\mathbf{c}}\|_\infty = \max_i |\widehat{c}_i|$. Ainsi, modulo une constante indépendante de l , on obtient le critère MDL

$$L(\mathbf{y}) = (m + l + 1)(\log_2(\|\widehat{\mathbf{c}}\|_\infty) + 1 - \log_2(\varepsilon)) + N \log_2(RMS) + l \log_2(N). \quad (6)$$

4.2. Influence de la précision de description de $\widehat{\mathbf{c}}$

Pour $\widehat{\mathbf{t}}$ donné, on étudie l'influence de la précision ε de description de $\widehat{\mathbf{c}}$ sur le modèle reconstruit. Soit $\widetilde{\mathbf{c}}$ une approximation de $\widehat{\mathbf{c}}$ telle que $\|\widetilde{\mathbf{c}} - \widehat{\mathbf{c}}\|_\infty \leq \varepsilon$. Selon le choix de l'évaluation de l'erreur entre les courbes estimée $\mathbf{B}\widehat{\mathbf{c}}$ et reconstruite $\mathbf{B}\widetilde{\mathbf{c}}$, on détermine une précision ε minimale.

Proposition 4.1 On a $\|\mathbf{B}\widetilde{\mathbf{c}} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty \leq \varepsilon$.

Preuve. D'après (1), on a $\|\mathbf{B}\|_\infty = 1$ et donc

$$\|\mathbf{B}\widetilde{\mathbf{c}} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty \leq \|\mathbf{B}\|_\infty \|\widetilde{\mathbf{c}} - \widehat{\mathbf{c}}\|_\infty \leq \varepsilon.$$

Ainsi, si l'on souhaite un écart maximal entre les points des courbes $\mathbf{B}\widehat{\mathbf{c}}$ et $\mathbf{B}\widetilde{\mathbf{c}}$ inférieur à l'écart entre les données et $\mathbf{B}\widehat{\mathbf{c}}$, il suffit de fixer $\varepsilon = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty$.

Corollaire 4.1 L'erreur RMS entre les courbes optimale $\mathbf{B}\widehat{\mathbf{c}}$ et reconstruite $\mathbf{B}\widetilde{\mathbf{c}}$ est inférieure à ε .

Preuve. On rappelle que pour tout $\mathbf{x} \in \mathbb{R}^N$, on a

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{N} \|\mathbf{x}\|_\infty \quad (7)$$

et la RMS entre $\mathbf{B}\widetilde{\mathbf{c}}$ et $\mathbf{B}\widehat{\mathbf{c}}$ vérifie, d'après la prop. 4.1,

$$\|\mathbf{B}\widetilde{\mathbf{c}} - \mathbf{B}\widehat{\mathbf{c}}\|_2 / \sqrt{N} \leq \|\mathbf{B}\widetilde{\mathbf{c}} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty \leq \varepsilon.$$

Ainsi, pour obtenir une erreur RMS entre les courbes $\mathbf{B}\widehat{\mathbf{c}}$ et $\mathbf{B}\widetilde{\mathbf{c}}$ inférieure à l'erreur RMS entre $\mathbf{y}$ et $\mathbf{B}\widehat{\mathbf{c}}$, il suffit de fixer $\varepsilon = RMS = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_2 / \sqrt{N}$.

4.3. Critères MDL pour les B-splines

Le critère MDL utilisé est défini par (6). On distingue deux versions selon la précision ε en la supposant fonction de l'erreur de reconstruction du modèle B-spline. On considère qu'il n'est pas nécessaire que l'erreur entre les courbes estimée et reconstruite soit inférieure à l'erreur d'estimation. D'après le paragraphe 4.2, on obtient les versions du critère : (a) : $\varepsilon = RMS$, et (b) : $\varepsilon = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty$.

5. PROTOCOLE EXPÉRIMENTAL

L'objectif est d'estimer le modèle B-spline $\widehat{\theta}^l$ et d'établir un compromis, à l'aide des critères MDL ci-dessus, entre sa qualité de modélisation (évaluée par l'erreur RMS) et son nombre de degrés de liberté (d.d.l.) l . On introduit les hypothèses méthodologiques communes à toutes les expériences avant de présenter dans le prochain paragraphe deux expériences permettant d'étudier : la relation entre l'erreur RMS et le nombre de d.d.l. et le comportement des différents critères MDL.

Les expériences sont réalisées sur 500 phrases (soit environ 7000 syllabes) choisies aléatoirement parmi un corpus d'environ 7000 phrases enregistrées. L'enregistrement a été réalisé dans un studio professionnel. Le signal acoustique a été annoté puis segmenté en unités acoustiques. La fréquence laryngienne moyenne, F_0 , a été analysée de manière automatique à l'aide de la fonction d'auto-corrélation. Ensuite, un algorithme a été appliqué à la chaîne d'unités phonétiques de manière à repérer chaque syllabe.

Pour l nœuds internes, le modèle $\widehat{\theta}^l$ possède $(2l + m + 1)$ paramètres et si ce nombre est supérieur au nombre N de valeurs de F_0 , il est plus économique de conserver ces dernières. Les courbes ayant un nombre différent d'observations, on normalise le nombre de d.d.l. l pour chaque courbe en le divisant par le nombre N d'observations. Cela est nécessaire pour comparer les moyennes de nombre de d.d.l. Un nombre de d.d.l. normalisé à 1 correspond alors à l'ensemble de ces points.

Toutes les expériences proposées font intervenir le calcul des valeurs moyennes de l'erreur RMS et du nombre de d.d.l. des modèles B-splines. Ces moyennes sont estimées par intervalles de confiance au niveau 99% établis selon une méthodologie bootstrap.

6. EXPÉRIENCES ET RÉSULTATS

Pour chaque courbe observée, le critère MDL sélectionne la structure du modèle parmi l'ensemble des structures de modèles possibles. La première étape consiste à estimer les modèles θ^l en faisant varier l . Ensuite, on applique les critères MDL proposés afin de sélectionner la

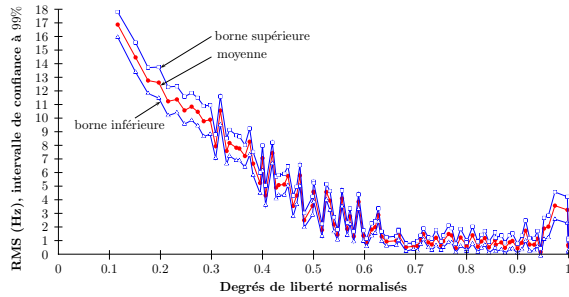


FIG. 1: Évolution des intervalles de confiance à 99% pour l'erreur RMS en fonction du nombre normalisé de d.d.l.

meilleure structure $\hat{l}$. On présente ci-après les différentes expériences réalisées pour évaluer ces critères.

6.1. Lien entre RMS et nombre de d.d.l.

La figure 1 permet de mesurer l'impact du nombre de paramètres sur la qualité d'estimation. Le nombre normalisé de d.d.l. présenté varie de 0 à 1. Toutes les courbes du corpus sont prises en compte. Lorsque le nombre de d.d.l. augmente, l'erreur RMS diminue. Ce résultat prévisible justifie la recherche d'un compromis entre la précision du modèle et sa complexité. Notamment, on remarque un point d'inflexion dans la courbe pour un nombre de d.d.l. normalisé moyen voisin de 0.65. Cette valeur correspond à une RMS moyenne proche de 1Hz. La remontée de l'erreur moyenne en fin de courbe montre une faiblesse du recuit simulé lorsque le nombre de nœuds devient important. De plus, la condition (4) n'est plus respectée. Quant aux fluctuations de la courbe, elles sont sans doute dues au caractère probabiliste de SA. Ainsi, malgré une optimisation globale par l'algorithme SA, la solution trouvée n'est pas forcément optimale. Le but du critère MDL étant de déterminer ce compromis, un critère satisfaisant devrait estimer une erreur RMS moyenne et un nombre de d.d.l. moyen aux environs du point d'inflexion de la courbe.

6.2. Analyse des critères MDL proposés

Les résultats de l'évaluation des critères MDL sur le corpus sont résumés dans le tableau 1 (intervalles de confiance à 99%). Deux modes variables ont été testés : l'utilisation de $\|y - \hat{B}\|_\infty$ améliore légèrement les résultats de RMS moyenne. En effet, d'après (7), $RMS \leq \|y - \hat{B}\|_\infty$. Le critère (b) est donc moins pénalisant et sélectionne un nombre supérieur de d.d.l. que (a). Si l'on privilégie l'erreur RMS, le critère (b) peut être qualifié de meilleur. De plus, le critère (b) se positionne dans la zone d'inflexion de la courbe au point (0.627, 0.42).

Dans [1] nous avons comparé un modèle B-spline et un modèle spline dans un cadre expérimental assez proche. Nous avons montré, tableau 1 page 4 de l'article, que des modèles B-splines avec un nombre normalisé de d.d.l. de l'ordre de 60% conduisent à une erreur RMS moyenne de l'ordre de 3Hz (on retrouve ce résultat figure 1 abscisse 0.6). Un modèle spline conduisait quant à lui à une erreur RMS moyenne de l'ordre de 12Hz. Dans [6], il a été montré que MoMel conduit à 6Hz d'erreur RMS en moyenne dans le meilleur des cas. Ces résultats suggèrent d'une part qu'un modèle B-spline est plus efficace qu'un modèle spline et d'autre part qu'un critère MDL à précision variable améliore la performance du modèle B-spline.

TAB. 1: Intervalles de confiance à 99% pour l'erreur RMS et le nombre normalisé de d.d.l.

Critères	d.d.l. norm.	RMS (Hz)
(a)	0.599 ± 0.006	0.68 ± 0.11
(b)	0.627 ± 0.006	0.42 ± 0.07

7. CONCLUSION

Dans cet article, nous avons présenté une nouvelle approche pour la modélisation de courbes de F_0 par l'utilisation d'un modèle B-spline. La précision d'une telle modélisation est importante pour caractériser la mélodie en traitement de la parole. Le modèle B-spline généralise le modèle spline et permet de décrire avec précision les irrégularités locales de la courbe. La principale contribution de cet article concerne l'estimation du nombre de paramètres libres du modèle B-spline grâce à une méthodologie MDL. Appliquée à la modélisation syllabique de contours de F_0 , cette approche conduit à une erreur RMS moyenne de 0.42Hz pour un nombre moyen de degré de liberté normalisé de 0.63 (un nombre de degré de liberté à 1 signifie qu'on utilise autant de paramètres que de points de la courbe). Ces valeurs de RMS sont d'une part inférieures aux seuils de JND pour le F_0 (de l'ordre de quelques Hz) et d'autre part obtenues avec un facteur de compression relativement important (en moyenne 37%).

RÉFÉRENCES

- [1] N. Barbot, O. Boëffard, and D. Lolive. F0 stylisation with a free-knot b-spline model and simulated-annealing optimization. In *Proc. Eurospeech Conf.*, pages 325–328, 2005.
- [2] M.A.F. Figueiredo, J.M.N. Leitão, and A.K. Jain. Unsupervised contour representation and estimation using b-splines and a minimum description length criterion. *IEEE Trans. Image Proc.*, 9(6) :1075–1086, 2000.
- [3] Mark H. Hansen and Bin Yu. Model selection and the principle of minimum description length. *J. Amer. Stat. Assoc.*, 96(454) :746–774, 2001.
- [4] D.J. Hirst, A. Di Cristo, and R. Espesser. Levels of representation and levels of analysis for the description of intonation systems. In *M. Horne (Ed.), Prosody : Theory and Experiment*, Kluwer Academic Publisher, 14 :51–87, 2000.
- [5] T.C.M. Lee. An introduction to coding theory and the two-part minimum description length principle. *Intl. Stat. Review*, 69(2) :169–183, 2001.
- [6] S. Mouline, O. Boëffard, and P.C. Bagshaw. Automatic adaptation of the momel f_0 stylisation algorithm to new corpora. In *Proc. of ICSLP*, 2004.
- [7] A. Raux and A.W. Black. A unit selection approach to f0 modeling and its application to emphasis. In *Proc. ASRU Conf.*, pages 700–703, 2003.
- [8] S. Sakai and J. Glass. Fundamental frequency modeling for corpus-based speech synthesis based on statistical learning techniques. In *Proc. ASRU Conf.*, pages 712–717, 2003.
- [9] P. Taylor. Analysis and synthesis of intonation using the tilt model. *J. Acoust. Soc. America*, 107 :1697–1714, 2000.