Melodic contour estimation with B-spline models using a MDL criterion

Damien Lolive, Nelly Barbot, Olivier Boeffard

IRISA / University of Rennes 1 - ENSSAT
6 rue de Kerampont, B.P. 80518, F-22305 Lannion Cedex
France
{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 knot number estimation is provided by MDL methodology. Three criteria are proposed: control points are first considered as integer values, next as real coefficients with fixed precision and then with variable precision. Experiments are conducted in a speech processing context on a 7000 syllables french corpus. We show that a variable precision criterion gives better results in terms of RMS error (0.42Hz) as well as in terms of reduction of the number of B-spline degrees of freedom (63% of the full model).

1. Introduction

Intonational speech models are widely used in the speech technology field. In particular, Text-to-Speech synthesis (TTS) cannot circumvent the use of such models to predict melodic contours from text and speaking styles. More recently, these models were used to find an optimal sequence of acoustic units taking account of prosodic characteristics [1]. Although intonation results from a combination of numerous linguistic factors, this article focuses on the acoustic parameters known to be the main factor of prosodic perception, which is the fundamental frequency F_0. F_0 contours, extracted from speech signal, ensue from the evolution of the vibration of the vocal folds over time. A wide range of publications deals with modelling of such contours. Especially, we can cite models like MoMel [2], Tilt [3], as well as Sakai and Glass's model [4] based on regular spline functions.

In this article, we propose a B-spline stylisation which integrates variable local regularity notion without affecting the global approximation of the F_0 curve, estimated according to the least square error criterion. In [5], a comparison between splines and B-splines ability to model F_0 is presented using a least squares criterion. The parameters of the B-spline model are the knot number and their placement. A free-knot placement allows to catch contour irregularities. A Monte-Carlo type algorithm (Simulated-Annealing) is implemented to circumvent the combinatorial complexity of the free-knot placement. Our main goal is then to estimate the optimal number of knots and to determine a parsimonious class of B-spline curves. The framework we choose is the minimum description length (MDL) criterion which offers an efficient compromise between model precision and its number of parameters.

In section 2, we introduce the B-spline model and its parameters, estimated in section 3 according to a least squares criterion. In section 4, we present the MDL criteria to optimize the number of parameters of the model. In section 5, the experimental protocol is described and results are given in section 6.

2. B-spline model

In this section, we present the spline functions and their generalization to B-spline curves which have the ability to model open curves with variable local regularity.

Let us consider an interval $[a, b]$ and its subdivision into $l + 1$ sub-intervals: $a = t_m < \ldots < t_{m+l+1} = b$. A spline of degree m associated to (t_i) is a polynomial function of degree m inside each sub-interval and belongs to the class of functions $\mathcal{C}^{m-1}$, which are $m - 1$ times continuously differentiable, on $[a, b]$. The l transition points are called the internal knots. The set of splines of degree m is a linear space of dimension $m + l + 1$.

In [2], the MoMel algorithm aims to provide melodic contour representation using a quadratic spline. For that, n target points are estimated to constitute the stationary points of a spline of degree 2. Transition points (t_i) correspond to the abscissas of the n target points and their $n - 1$ median points. Thus, the quadratic spline is associated to $l = 2n - 3$ internal knots and is characterized by the $2n$ constraints due to the n points to interpolate with an horizontal tangent.

The linear space of splines associated to (t_i) admits a basis of B-spline functions. In order to define them, we add to this knot sequence the $(m+1)$ times duplicated values $t_0 = t_m = a$ and $t_{m+l+1} = t_{2m+l+1} = b$, called external knots. We denote $\mathbf{t}$ the vector containing the internal and external knots. Then, B-splines are recursively defined: for i from 0 to $2m + l$, the B-splines of degree 0 are given by

$$B_0^i(t) = \mathbf{1}_{[t_i, t_{i+1})}(t)$$

and for i from 0 to $m + l$, the B-splines of degree m are given by

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

where quotients equal zero if $t_i = t_{i+m}$ or $t_{i+1} = t_{i+m+1}$. Moreover, B-splines are non-negative and satisfy

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

Consequently, a spline function of degree m can be written as a linear combination of B-spline functions of same degree. Spline generalization to B-spline curve allows duplicated knots, the knot multiplicity corresponding to the number of times it appears in $\mathbf{t}$. This way, a B-spline curve of degree m associated to a knot vector $\mathbf{t}$ can be written as a linear combination of B-spline functions of degree m where the coefficients $c_0, \ldots, c_{m+l}$ are called control points.

Let a sequence $x_0, \ldots, x_{N-1}$ inside $[a, b]$, the matrix representation of the B-spline curve at point x_j is given by

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

where $B_m^i(x_j)$ is the (j, i)-th entry of the matrix $\mathbf{B}$, $\mathbf{c}$ is the column vector containing the control points and

$$\theta^l = (t_{m+1}, \ldots, t_{m+l}, c_0, \ldots, c_{l+m}) .$$

Now we assess the impact of each parameter of a B-spline curve g_{θ^l}. The knot effect on the curve g_{θ^l} is mainly controlled by its multiplicity. If we consider an internal knot t_i, greater its multiplicity m_i is, lower is the smoothness of the curve at t_i. More precisely, if g_{θ^l} belongs to $\mathcal{C}^{m-1}$ between two consecutive knots, it belongs to $\mathcal{C}^{m-m_i}$ at knot t_i. For instance, if $m_i = m - 1$ the curvature of the B-spline curve changes at t_i, if $m_i = m$ the curve can not be differentiated at t_i and if $m_i > m$, it is not continuous at t_i. As regards control points, they have a local influence. Indeed, suppose we change the position of c_i, it only alters the term $c_i B_m^i$ in (2), and since B_m^i is zero outside $[t_i, t_{i+m+1})$, it does not propagate outside the corresponding curve segment.

3. Parameter estimation

Let a set of measurements $\{(x_j, y_j), 0 \leq j \leq N - 1\}$ of a melodic contour, where (x_j) is a non-decreasing sequence. Take the problem of the B-spline curve g_{θ^l} of degree m such as the points $(x_j, g_{\theta^l}(x_j))$ best fit these data using the least squares error criterion. In this section, we suppose the number l of internal knots known, and let $t_0 = t_m = x_0$ and $t_{m+l+1} = t_{2m+l+1} = x_{N-1}$. First, we derive the optimal control points and then we estimate the internal knot location.

3.1. Control points

In this section, the knot vector $\mathbf{t}$ is supposed known and we determine the B-spline curve g_{θ^l} of degree m, that is to say its control points, which minimizes the mean square error with the data. Let us denote $\mathbf{y}$ the column vector containing the y_j, the control points are estimated such as

$$\widehat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{y} - \mathbf{Bc}\|_2^2$$

according to (2). After derivation, we obtain

$$\widehat{\mathbf{c}} = \left(\mathbf{B}^T \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{y} \qquad (3)$$

if the matrix $\mathbf{B}^T \mathbf{B}$ is invertible, i.e. when the columns of $\mathbf{B}$ are linearly independent. That is to say $\mathbf{t}$ must not have knots with multiplicity order greater than $m+1$ to avoid null columns of $\mathbf{B}$. Besides, the number N of lines of $\mathbf{B}$ has to be really greater than the number of columns in order to differentiate the columns, i.e.

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

Finally, for a given knot sequence $\mathbf{t}$, the optimal B-spline curve of degree m according to a least squares criterion is described by the N following points

$$x_j \longrightarrow (\mathbf{B}\widehat{\mathbf{c}})_j = \left(\mathbf{B} \left(\mathbf{B}^T \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{y}\right)_j . \qquad (5)$$

3.2. Knot placement

Let us now consider the optimization of the internal knot location. We introduce the maximum likelihood criterion and we choose a Monte-Carlo strategy (Simulated-Annealing, SA) to compute the solution $\widehat{\mathbf{t}}$ of this NP-hard problem. We restrict the knot location to the set of observations, and knots can then be represented by integers between 0 and $N - 1$. Our experiments show that such a restriction does not seem to decrease the quality of the final curve estimate.

Let us denote e_j the error between the observation y_j at time x_j and its estimation $g_{\theta^l}(x_j)$. To simplify calculus, the errors $e_0, \ldots, e_{N-1}$ are supposed independent and identically distributed according to a zero mean Gaussian law with variance σ^2. Then, given the model, the log-likelihood of $\mathbf{y}$ is

$$\log p\left(\mathbf{y}; \theta^l\right) = -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Bc}\|_2^2 - \frac{N}{2} \log\left(2\pi\sigma^2\right) . \quad (6)$$

Let us notice that $\widehat{\mathbf{c}}$, defined by (3), is the maximum likelihood estimate of $\mathbf{c}$. The maximum likelihood estimate $\widehat{\mathbf{t}}$ of $\mathbf{t}$ is defined by

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

where the matrix $\mathbf{B}$ depends on $\mathbf{t}$.

To determine $\widehat{\mathbf{t}}$, we used a global optimization algorithm based on the Simulated-Annealing (SA) strategy, presented in [5]. The proposed relation between the parameters of the model and the random distributions of the SA algorithm is as follows:

1. SA samples a vector $\mathbf{v}$ inside $\{x_1, \ldots, x_{N-2}\}^l$.

2. An internal knot vector $\mathbf{t}^*$ is defined by sorting the coordinates of $\mathbf{v}$.

3. A knot vector $\mathbf{t}$ is defined by adding $(m + 1)$ times x_0 and x_{N-1} to the extremities of $\mathbf{t}^*$.

4. If two consecutive knots lie in an interval smaller than 5% of the full range $[x_0, x_{N-1}]$, the two knots are merged and the multiplicity order of the first knot is increased by one. This process is then applied recursively to all the knots defining $\mathbf{t}$. Let us recall that knot multiplicity must not be greater than $m + 1$.

Before tackling the optimization of the number l of internal knots, it is useful to sum up the main choices and strategies as regards the considered B-spline model $\widehat{\theta^l}$. First, for a given l, we start to estimate an optimal knot-sequence $\widehat{\mathbf{t}}$ using the SA algorithm. Then with the associated matrix $\mathbf{B}$ and (3), we compute the optimal $\widehat{\mathbf{c}}$. In the following section, we look for the best model, that is to say the optimal value of l, according to a MDL criterion.

4. B-spline and MDL

The minimum description length (MDL) criterion enables a compromise between the quality of a model $\widehat{\theta^l}$ and its complexity l. More precisely, MDL consists in determining $\widehat{l}$ which

minimizes the description length $L(\mathbf{y})$ of the data $\mathbf{y}$. According to [6], we have

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

where $L\left(\widehat{\theta^l}\right)$ and $-\log_2 p\left(\mathbf{y}; \widehat{\theta^l}\right)$ stand respectively for the description length of the estimated model and the observations conditioned by the model. The expression (6), using the unknown error variance σ^2, we use its maximum likelihood estimate $\widehat{\sigma^2}$:

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

Hence, σ is estimated by the root mean square error (RMS) between the observations and the B-spline curve. Thus, using expression (6), we have

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

4.1. Principle of solution

We consider now the description length of the vector $\widehat{\theta^l}$. This one is composed of l internal knots and $(l + m + 1)$ control points. We suppose that the knots have all the same description length, as well as the control points.

Regarding the internal knots, they are inside the interval (x_0, x_{N-1}) at observation places. Hence, each knot can be represented by an integer between 0 and $N - 1$ and so $\log_2(N)$ bits are sufficient to encode it.

As for the control points, they are real parameters estimated from N data points. When N is large, the length of a real parameter is generally approached by $\log_2(\sqrt{N})$ which is asymptotically optimal [6]. However, each control point has a local influence on the B-spline model $g_{\widehat{\theta^l}}$ and if the number of observations is small the asymptotic approximation is not well adapted [7]. We then consider a description length of a control point based on a uniform prior on the interval $[-\alpha, \alpha]$ [8]. For a given control point description length precision ε, its length is given by $\log_2(\alpha) + 1 - \log_2(\varepsilon)$. The choices of α and ε are discussed in the following paragraph. In short, the description length of $\widehat{\theta^l}$ can be written as

$$L\left(\widehat{\theta^l}\right) = l\log_2(N) + (m+l+1)\left(\log_2(\alpha) + 1 - \log_2(\varepsilon)\right) .$$

4.2. Theoretical bounds on control points

For a given $\widehat{\mathbf{t}}$, the control points $\widehat{\mathbf{c}}$ are assumed generated by an uniform prior on $[-\alpha, \alpha]$. The maximum likelihood estimate of α is $\|\widehat{\mathbf{c}}\|_\infty = \max_i |\widehat{c}_i|$. Thus, modulo a constant independent of l, we obtain the first MDL criterion

$$\begin{aligned} L(\mathbf{y}) = \ & (m+l+1)\left(\log_2\left(\|\widehat{\mathbf{c}}\|_\infty\right) + 1 - \log_2(\varepsilon)\right) \\ & + N\log_2(RMS) + l\log_2(N) \end{aligned} \quad (7)$$

denoted ***criterion (a)***.

Considering the expression (3) of $\widehat{\mathbf{c}}$, we establish a bound on control point definition interval and we can then consider the associated uniform prior.

Proposition 4.1 *Let us denote μ the smallest singular value of the matrix $\mathbf{B}$, we have $\mu > 0$ and $\|\widehat{\mathbf{c}}\|_\infty \leq \|\mathbf{y}\|_2/\mu$.*

Proof. The matrix $\mathbf{B}$ being of maximal rank, i.e. $m + l + 1$, its smallest singular value μ is positive. According to [9],

$$\left\|(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\right\|_2 = 1/\mu$$

and using (3), we have

$$\begin{aligned} \|\widehat{\mathbf{c}}\|_\infty \ &\leq \ \|\widehat{\mathbf{c}}\|_2 = \left\|(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}\right\|_2 \\ &\leq \ \left\|(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\right\|_2 \|\mathbf{y}\|_2 \\ &\leq \ \|\mathbf{y}\|_2/\mu . \end{aligned}$$

From proposition 4.1, if the control points are assumed uniformly distributed on $[-\|\mathbf{y}\|_2/\mu, \|\mathbf{y}\|_2/\mu]$, we obtain the following MDL criterion

$$\begin{aligned} L(\mathbf{y}) = \ & (m+l+1)\left(\log_2\left(\|\mathbf{y}\|_2/\mu\right) + 1 - \log_2(\varepsilon)\right) \\ & + N\log_2(RMS) + l\log_2(N) \end{aligned} \quad (8)$$

denoted ***criterion (b)***.

4.3. Control point precision

For a given $\widehat{\mathbf{t}}$, we now study the influence of the control point precision ε on the rebuilt model. Let us note $\widetilde{\mathbf{c}}$ an approximation of $\widehat{\mathbf{c}}$ such that $\|\widetilde{\mathbf{c}} - \widehat{\mathbf{c}}\|_\infty \leq \varepsilon$. According to the chosen error evaluation between the estimated curve $\mathbf{B}\widehat{\mathbf{c}}$ and the rebuilt one $\mathbf{B}\widetilde{\mathbf{c}}$, we derive a sufficient precision ε.

Proposition 4.2 *We have $\|\mathbf{B}\widetilde{\mathbf{c}} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty \leq \varepsilon$.*

Proof. According to (1), $\|\mathbf{B}\|_\infty = 1$ and we get

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

As a result, if we want a variation between the estimated and rebuilt points smaller than the error between the data and the estimated points, choosing a control point precision of $\varepsilon = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty$ is sufficient.

Corollary 4.1 *The RMS error between the rebuilt curve $\mathbf{B}\widetilde{\mathbf{c}}$ and the optimal one $\mathbf{B}\widehat{\mathbf{c}}$ is less than ε.*

Proof. We recall that for any vector $\mathbf{x}$ of $\mathbb{R}^N$,

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{N}\|\mathbf{x}\|_\infty \quad (9)$$

and using proposition 4.2, the RMS error between $\mathbf{B}\widetilde{\mathbf{c}}$ and $\mathbf{B}\widehat{\mathbf{c}}$ then verifies

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

Thus, if we wish a RMS error between the estimated B-spline curve $\mathbf{B}\widehat{\mathbf{c}}$ and $\mathbf{B}\widetilde{\mathbf{c}}$ less than the RMS error between $\mathbf{y}$ and $\mathbf{B}\widehat{\mathbf{c}}$, taking $\varepsilon = RMS = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_2 / \sqrt{N}$ is sufficient.

4.4. MDL criteria for B-splines

In paragraph 4.2, we have introduced two MDL criteria depending on the choice of the control point description length, called (a) and (b), defined by (7) and (8). For each of them, we distinguish three sub-criteria in function of the precision ε. The first one considers ε as fixed and the others as a variable precision depending on the model. We consider that the error between the estimated and rebuilt curves has not to be lower than the one between the data and the estimated curve. According to paragraph 4.3, we choose $\varepsilon = RMS$ and then $\varepsilon = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty$. From criteria (a) and (b), we obtain three sub-criteria:

- (a).1 and (b).1: ε fixed for all curves,
- (a).2 and (b).2: $\varepsilon = RMS$,
- (a).3 and (b).3: $\varepsilon = \|\mathbf{y} - \mathbf{B}\widehat{\mathbf{c}}\|_\infty$.

5. Experimental protocol

The main objective consists in estimating the B-spline model $\widehat{\theta^l}$ using the above MDL criteria in order to obtain a compromise between the model quality (assessed by the RMS error) and the number l of degrees of freedom (D.F.). We introduce the methodological hypotheses common to all experiments which answer to the three following questions: what is the relation between the RMS error and the number of D.F. of the model ? What is the MDL criteria sensibility in function of the precision ε ? Finally, how do the proposed criteria behave ?

The speech corpus used in these experiments is a French one made of approximately 7000 syllables randomly extracted from a 7000 sentences corpus. The acoustic signal was recorded in a professional recording studio; the speaker was asked to read the text. Then, the acoustic signal was annotated and segmented into phonetic units. The fundamental frequency, F_0, was analyzed in an automatic way according to an estimation process based primarily on the autocorrelation function of the speech signal. Next, an automatic algorithm was applied to the phonetic chain pronounced by the speaker so as to find the underlying syllables. We choose the syllable as the minimal support of a melodic contour. Our objective is to measure the performance of the unsupervised stylisation of melodic contours using different decision criteria.

For l internal knots, the model $\widehat{\theta^l}$ depends on $(2l + m + 1)$ parameters, and if this number is greater than the number N of F_0 values, it is more economical to memorize the curve. When we estimate a B-spline model, we then choose to satisfy the following condition on the internal knot number l: $N \geq (2l + m + 1)$. The curves having different lengths, we normalize the number of internal knots in order to calculate the means of comparable number of D.F. for all the curves. More precisely, we divide the knot number of a B-spline model by the number of points of the curve. A normalized number of D.F. equal to 1 will then correspond to the whole curve, i.e. to a full model.

All proposed experiments use mean values calculation: the mean value of the RMS errors and the mean value of the B-spline models D.F. These ones are formulated as confidence intervals with a confidence threshold equal to 99%. Considering the nature of our experiments, a single experimental sample containing 7000 curves is used, the confidence intervals are obtained by resampling the empirical distributions (Bootstrap methodology).

6. Results and discussion

The MDL criterion selects the model structure between all the possible models for a given syllable. So the first stage is the estimation of the models θ^l for each syllable varying internal knot number l. The second stage is to apply the MDL criteria on each syllable to decide which model $\widehat{\theta^l}$ is the best one. In this section, we present the different experiments we carried out.

6.1. RMS error versus D.F.

These experiments are conducted in order to observe the evolution of the RMS error in function of the degrees of freedom number. The obtained curve (figure 1) permits to measure the parameter number impact on the estimated models quality. The normalized D.F. number varies from 0 to 1. All syllables of the corpus are taken into account in this figure.

When D.F. number increases, RMS error decreases. Although this result was foreseeable, it justifies the search of a
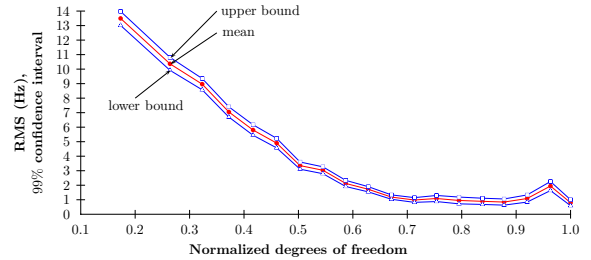

Figure 1: RMS error 99% confidence interval evolution in function of the normalized degrees of freedom.

compromise between the model precision and its complexity. In particular, we notice a point of inflection in the curve with an average normalized D.F. number close to 0.65. This value corresponds to a mean value of RMS error close to 1Hz.

The increase at the end of the curve probably shows a weakness of the SA strategy when the knot number becomes high. Indeed, the condition (4) is no more satisfied. Although the optimization process is global, SA does not necessarily provide an optimal knot vector. Finally, the MDL goal being to obtain a compromise between precision and complexity, a satisfying criterion should estimate a mean RMS error and a mean normalized D.F. number near the inflection point of the curve.

6.2. MDL and ε relation

It is important to compare variable precision MDL criteria to fixed precision ones. For that, we study the evolution of the RMS error according to all the possible values of ε (criteria (a).1 et (b).1). Thus, we assess the influence of the parameter precision ε on the RMS error and the D.F. number.

Figure 2 represents the RMS error confidence intervals in function of ε values. Figure 3 shows the evolution of the normalized D.F. number in function of ε. RMS error values and ε values are represented with a logarithmic scale. Notice that the fact of considering the control points as integers corresponds to the choice of $\varepsilon = 1$.

6.2.1. Criterion (a)

Less precisely the model is coded (ε larger), higher knot number is allowed. Indeed, when ε increases, the term $-\log \varepsilon$ decreases, the criterion selects a higher value of l and the RMS error then decreases. In figure 2, we can notice a quite significant variation of the RMS error. Indeed, there is a scale factor close to 10 between its maximum and minimum values. Moreover, we observe an important increase of the D.F. number. To sum up, the curve variations underline the impact of ε on the RMS error and the chosen D.F. number.

6.2.2. Criterion (b)

The same observations as for criterion (a) are valid for criterion (b). Indeed, the influence of ε is important, it governs the criterion performance. However, this criterion gives higher RMS error values than the previous one. According to proposition 4.1, the control point description length is greater and penalizes more the MDL criterion (b). Therefore, the average D.F. number stemming from (b) is lower than the one from (a), inducing a RMS error increase. This observation is coherent with the evolution of the RMS error in function of the D.F. number in figure 1.
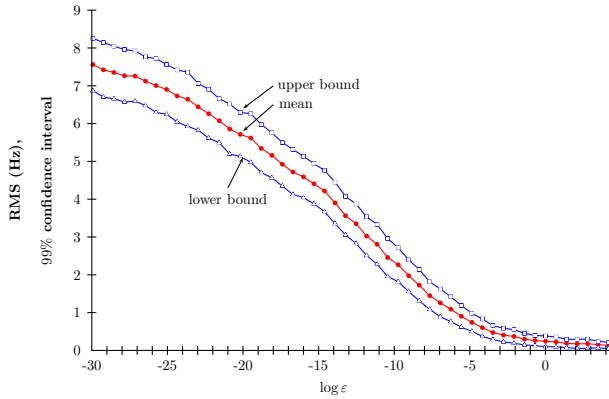
Figure 2: RMS error 99% confidence interval in function of $\log \varepsilon$ for criterion (a).
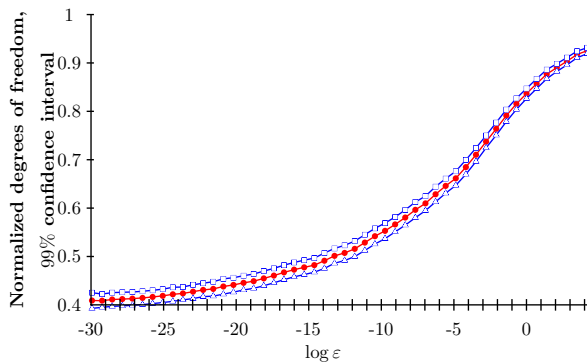

Figure 3: Normalized degrees of freedom 99% confidence interval evolution in function of $\log \varepsilon$ for criterion (a).

6.3. Proposed MDL criteria analysis

To evaluate the proposed criteria, we calculate confidence intervals on the RMS error and the normalized D.F. number for the MDL selected models. The results are summarized in table 1. By comparing these confidence intervals to previous results, we will be able to observe the reliability of the proposed criteria. Contrary to the preceding experiments, ε is variable (criteria (a).2, (a).3, (b).3 et (b).3).

These results permit the distinction of two compromises. First, the criterion (a).2 gives a mean RMS error of 0.68Hz with a normalized D.F. number of 0.599, while criterion (b).2 gives respectively 1.57Hz and 0.544. Criterion (b) uses a control point description length higher than the one used in (a). Therefore, it penalizes more and the selected l values are smaller. The mean RMS error then increases.

Secondly, two variable modes were tested for each criterion. The use of $\|\mathbf{y} - \mathbf{B}\hat{\mathbf{c}}\|_\infty$ slightly improves the RMS error results for each criterion and implies an increase of the D.F. number. Indeed, according to (9), $RMS \leq \|\mathbf{y} - \mathbf{B}\hat{\mathbf{c}}\|_\infty$ so the criteria (a).3 et (b).3 are less penalizing. If the RMS error is privileged, the third versions of the criteria are better than the others.

To compare fixed and variable precision criterion, it is necessary to compare the normalized D.F. numbers, taking equal mean RMS errors. Similarly, a reciprocal comparison is essential. For criterion (a).3, if we choose a mean RMS error equal to 0.42Hz, figure 2 gives the $\log \varepsilon$ value -3. By reporting this

Table 1: 99% confidence intervals for the RMS error and the Normalized Degrees of Freedom (Norm. D.F.).

Criteria	RMS (Hz)	norm. D.F.
(a).2	0.68 ± 0.11	0.599 ± 0.006
(a).3	0.42 ± 0.07	0.627 ± 0.006
(b).2	1.57 ± 0.16	0.544 ± 0.006
(b).3	1.11 ± 0.12	0.567 ± 0.006

value on figure 3, the obtained D.F. number is approximately 0.75. If we apply this reasoning in the same way with equal D.F., we can conclude that criterion (a).3 is better than a fixed precision one. The same reasoning is valid for (b).3.

We have to compare the variable precision criteria to the general evolution of the RMS error in function of the normalized degrees of freedom (fig.1). Criterion (a).3 is located near the inflection point of the curve at coordinates $(0.627, 0.42)$. Criterion (b).3 is before the inflection point at $(0.567, 1.11)$. We conclude that criterion (b) is less efficient than (a).

In [5], we compared a B-spline model to a spline model using an experimental framework close to the one used here. We showed, table 1 page 4 of the article, that B-spline models with a normalized degree order of 60% leads to a mean RMS error about 3Hz (this result can be found figure 1 abscissa 0.6). As for the spline model, it led to a mean RMS error around 12Hz. In [10], it has been shown that MoMel leads to 6Hz of mean RMS error in the best case. These results suggest, on the one hand, that a B-spline model outperforms a spline model and, on the other hand, a MDL criterion with variable precision improves B-spline model performance.

7. Conclusion

In this article, we present a new approach to estimate melodic contours using a B-Spline model. The precision of such a modelling is important to characterize melody in speech processing. B-spline model generalizes spline model and allows to describe precisely the local irregularities of the curve. The main contribution of this article concerns the estimation of an optimal knot number for the B-spline model thanks to the MDL methodology. Applied to F_0 contours modelling at the syllabic level, this approach leads to a mean RMS error of 0.42Hz with a normalized number of degrees of freedom equal to 0.63 (a normalized number of degrees of freedom equal to 1 corresponds to the use of as much parameters as points of the curve). These RMS values are, on the one hand, less than the F_0 JND threshold (around a few Hertz) and on the other hand, they are obtained with a compression factor relatively high (37% on average).

8. References

[1] A. Raux and A. Black, "A unit selection approach to f0 modeling and its application to emphasis," in *Proc. ASRU Conf.*, 2003, pp. 700–703.

[2] D. Hirst, A. D. 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 Pusblisher*, vol. 14, pp. 51–87, 2000.

[3] P. Taylor, "Analysis and synthesis of intonation using the

tilt model," *J. Acoust. Soc. America*, vol. 107, pp. 1697–1714, 2000.

[4] S. Sakai and J. Glass, "Fundamental frequency modeling for corpus-based speech synthesis based on statistical learning techniques," in *Proc. ASRU Conf.*, 2003, pp. 712–717.

[5] N. Barbot, O. Boeffard, and D. Lolive, "F0 stylisation with a free-knot b-spline model and simulated-annealing optimization," in *Proc. Eurospeech Conf.*, 2005, pp. 325–328.

[6] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 746–774, 2001.

[7] M. Figueiredo, J. Leitao, and A. Jain, "Unsupervised contour representation and estimation using b-splines and a minimum description length criterion," *IEEE Trans. Image Proc.*, vol. 9, no. 6, pp. 1075–1086, 2000.

[8] T. C. Lee, "An introduction to coding theory and the two-part minimum description length principle," *Intl. Stat. Review*, vol. 69, no. 2, pp. 169–183, 2001.

[9] G. Golub and C. V. Loan, *Matrix Computations*. Johns Hopkins Univ. Press, Baltimore, MA., 1989.

[10] S. Mouline, O. Boeffard, and P. Bagshaw, "Automatic adaptation of the momel f_0 stylisation algorithm to new corpora," in *Proc. of ICSLP*, 2004.