Pitch contour stylization using an optimal piecewise polynomial approximation[1]

Prasanta Kumar Ghosh and Shrikanth S. Narayanan*

Signal Analysis and Interpretation Laboratory, Department of Electrical Engineering,

University of Southern California, Los Angeles, CA 90089

prasantg@usc.edu, shri@sipi.usc.edu

Ph: (213) 821-2433, Fax: (213) 740-4651

Abstract: We propose a dynamic programming (DP) based piecewise polynomial approximation of discrete data such that the L_2 norm of the approximation error is minimized. We apply this technique for the stylization of speech pitch contour. Objective evaluation verifies that the DP based technique indeed yields minimum mean square error (MSE) compared to other approximation methods. Subjective evaluation reveals that the quality of the synthesized speech using stylized pitch contour obtained by the DP method is almost identical to that of the original speech.

EDICS: SPE-ANAL

Index Terms: pitch stylization, piecewise polynomial approximation.

1 Introduction

Piecewise approximation of data using polynomial functions of finite order is a problem of interest in many fields of science and engineering, such as compression of ECG signals [1], environment compensation in automatic speech recognition [2], design of embedded systems without floating point capabilities [3], and stylization of pitch contour [4]. Let $\{x_n\}_{n=1}^N$ be N data points. The piecewise polynomial approximation problem requires that K piecewise polynomial functions of order P have to be used to approximate $\{x_n\}_{n=1}^N$. In this paper, we derive an $O(KN^2)$ algorithm

based on dynamic programming (DP) which minimizes the L_2 norm of the approximation error. This algorithm finds the boundaries of piecewise segments and also the polynomial coefficients in each segment. We apply the proposed algorithm for pitch contour stylization, a key potential ingredient for many speech processing applications such as synthesis and speech understanding.

There have been many works in the literature on piecewise polynomial (in particular, linear) approximation of a function or data. Cantoni [9] proposed an optimal curve fitting technique with piecewise linear functions, but the solution requires explicit prespecified end points of segments. Tomek [10] proposed two heuristic algorithms for piecewise linear continuous approximation of functions of one variable but these were not formulated as an explicit optimization. More recently, Miroslav et al. [11] proposed a recursive formula for piecewise polynomial approximation of discrete functions. The proposed recursion finds the best polynomial fit to a set of local data points, which have to be specified explicitly. Thus this recursion does not provide optimal piecewise segment boundaries given a set of data points. Obata et al. [12] used fluency theory to obtain piecewise polynomial approximation of given data. However, for better approximation, adaptation of the class of fluency vector is required depending on the input data, although the authors in [12] do not address specific adaptation solutions.

A number of approximation algorithms have been proposed in the literature [4, 5, 6, 7, 8] for pitch stylization, but most are predominantly heuristic and are not formulated to directly optimize any specific objective metric. Optimality in terms of some objective function is necessary to understand the effect of parameterization of the pitch contour in a systematic way. In this regard, our proposed algorithm provides the flexibility of obtaining an optimal approximation for given choices of K and P. This offers the advantage of studying and comparing various possible piecewise parameterizations of the pitch contour in both objective and subjective manner.

It should be noted that Ranveig et al. [1] also proposed an $O(KN^2)$ algorithm using a directed graph (DG) approach, but we will show that the proposed algorithm gives a lower L_2 norm of the approximation error compared to that in [1]. The main difference between our

approach and that of [1] is that our approach does not assume the boundary points of the piecewise approximation to be some of the given data points while Ranveig et al [1] do have such constraints. This leads to a lower L_2 norm of the approximation error in our approach.

2 Problem Definition

Let $\{x_n\}_{n=1}^N$ be N pitch values (in general, any set of discrete data points), where n is the index variable of the data. The problem is to approximate $\{x_n\}_{n=1}^N$ using K piecewise polynomial functions of order P. This means we have to find out $K-1$ boundary indices $\{\eta_k\}_{k=1}^{K-1}$ for K piecewise segments and the polynomial coefficients $\left\{a_l^k\right\}_{l=0}^P, \quad k=1,...,K$ for all K piecewise segments. However, we need to first define a criterion with respect to which the $\{\eta_k\}$ and $\left\{a_l^k\right\}$ will be optimal.

Let the data $\{x_n\}_{n=1}^N$ be modeled as realizations of the random variables:

$$X_n = \sum_{l=0}^{P} a_l^k n^l + \varepsilon_n \quad , \quad \eta_{k-1} \leq n \leq \eta_k, \ k=1,...,K \tag{1}$$

where $\eta_0 = 1$ and $\eta_K = N$ and ε_n is independent identically distributed (i.i.d.) random variables, having normal distribution with mean 0 and variance σ^2 with corresponding probability density function (pdf) of ε_n being $f_{\varepsilon_n}(y) = \mathcal{N}(0,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}}, \quad -\infty \leq y \leq \infty$.

Thus, X_n $(\eta_{k-1} \leq n \leq \eta_k)$ are independent random variables with the pdf $f_{X_n}(x) = \mathcal{N}\left(\sum_{l=0}^P a_l^k n^l, \sigma^2\right)$. $\{\eta_k\}$ and $\left\{a_l^k\right\}$ are determined by maximizing the likelihood of the observation sequence given by $f_{X_1,...,X_N}(x_1,...,x_N) = \prod_{n=1}^N f_{X_n}(x_n)$ (because X_n are independent). It's easy to show that maximizing $f_{X_1,...,X_N}(x_1,...,x_N)$ is equivalent to minimizing $\left[\sum_{k=1}^K \sum_{n=\eta_{k-1}}^{\eta_k - sign(K-k)} \left(x_n - \sum_{l=0}^P a_l^k n^l\right)^2\right]$, which is the L_2 norm square of the residual ε_n. $sign(K-k) = 1$, when $k < K$. This is done to avoid counting boundary points twice. Thus, the opti-

mization problem becomes

$$\{\tilde{\eta}_k\},\ \left\{\tilde{a}_l^k\right\} = \underset{\{\eta_k\},\ \{a_l^k\}}{\arg\min} \left[\sum_{k=1}^{K} \sum_{n=\eta_{k-1}}^{\eta_k - sign(K-k)} \left(x_n - \sum_{l=0}^{P} a_l^k n^l\right)^2\right],\ k = 1, ..., K,\ l = 0, ..., P \quad (2)$$

$$\text{subject to } \sum_{l=0}^{P} a_l^{k-1} \eta_k^l = \sum_{l=0}^{P} a_l^k \eta_k^l,\quad k = 1, ..., K-1$$

The constraints ensure continuity at the boundaries of piecewise segments. This cost function is not differentiable w.r.t. η_k although it is differentiable w.r.t. a_l^k. A full search for optimal $\{\eta_k\}$ has an order complexity $O(N^{K-1})$. Instead, we derive a dynamic programming (DP) based solution which has an order complexity $O(N^2 K)$.

Also note that for $K = 1$ and $P = 1$, this problem becomes a simple least square problem [13]. Thus, the problem addressed here is a generalization of the least square approximation.

3 Dynamic programming (DP) based solution

DP works on the principle of doing locally best to achieve a globally best solution. Hence we first need to derive a solution of the following problem, which provides a best polynomial approximation (of known order P) of local data points $\{x_m\}_{m=M_1}^{M_2}$ such that the following L_2 norm of the approximation error is minimized subject to an initial constraint $\sum_{p=0}^{P} \alpha_p M_1^p = \beta$.

$$\{\tilde{\alpha}_p\} = \underset{\{\alpha_p\}}{\arg\min} \sum_{m=M_1}^{M_2} \frac{1}{2} \left(x_m - \sum_{p=0}^{P} \alpha_p m^p\right)^2,\ p = 0, ..., P \quad (3)$$

where P and β are known. This can be easily solved by the Lagrange multiplier method:

$$\text{Let } J(\{\alpha_p\}, \lambda) = \frac{1}{2} \sum_{m=M_1}^{M_2} \left(x_m - \sum_{p=0}^{P} \alpha_p m^p\right)^2 + \lambda \left(\beta - \sum_{p=0}^{P} \alpha_p M_1^p\right) \quad (4)$$

where λ is the Lagrange multiplier. Eqn. (4) has $P+2$ unknowns $\{\alpha_p\}_{p=0}^{P}$ and λ, which can be solved from the $P+2$ linear equations: $\left\{\frac{\partial J}{\partial \alpha_q}=0\right\}_{q=0}^{P}$ and $\left\{\frac{\partial J}{\partial \lambda}=0\right\}$. This can be written as

$$A\underline{\theta} = \underline{b} \tag{5}$$

where

$$A = \begin{pmatrix} \sum_{m=M_1}^{M_2} m^0 & ... & \sum_{m=M_1}^{M_2} m^P & M_1^0 \\ . & ... & . & . \\ . & ... & . & . \\ \sum_{m=M_1}^{M_2} m^P & ... & \sum_{m=M_1}^{M_2} m^{2P} & M_1^P \\ M_1^0 & ... & M_1^P & 0 \end{pmatrix}, \underline{\theta} = \begin{pmatrix} \alpha_0 \\ . \\ . \\ \alpha_P \\ \lambda \end{pmatrix} \text{ and } \underline{b} = \begin{pmatrix} \sum_{m=M_1}^{M_2} x_m \\ . \\ . \\ \sum_{m=M_1}^{M_2} x_m m^P \\ \beta \end{pmatrix}$$

When the number of data points is at least greater than the order of the polynomial, i.e., when $M_2 - M_1 \geq P$, $\underline{\theta}$ (and hence $\{\tilde{\alpha}_p\}$) can be obtained as follows:

$$\underline{\theta} = A^{-1}\underline{b} \tag{6}$$

Therefore, if two index points M_1 and $M_2 (\geq P + M_1)$ and a constraint value β are provided for the optimization problem in eqn. (3), eqn. (6) provides $\{\tilde{\alpha}_p\}_{p=0}^{P}$. Let $\delta(M_1, M_2, \beta)$ denote the corresponding Sum of Squared approximation Error (SSE), i.e.,

$$\delta(M_1, M_2, \beta) = \sum_{m=M_1}^{M_2} \left(x_m - \sum_{p=0}^{P} \tilde{\alpha}_p m^p \right)^2 \tag{7}$$

where $\sum_{p=0}^{P} \tilde{\alpha}_p M_1^p = \beta$. Let us now define the necessary terminologies for deriving the optimal solution of eqn. (2) using DP. Let $D_k(r)$ be the SSE of the approximation error for fitting k optimum polynomial functions of order P between x_1 and x_r $(kP + 1 \leq r \leq N)$[2]. Let $\xi_k(r)$ be

[2]The minimum value of r for fitting k polynomials of order P between x_1 and x_r is $kP + 1$. For example, considering $k = 1$, at minimum r should be $P + 1$ because we need a minimum of $P + 1$ data points between x_1

the backtracking pointer, that stores the starting index point for the k^{th} polynomial function for fitting k optimum polynomial functions of order P between x_1 and x_r ($kP+1 \leq r \leq N$). Let $\beta_k(r)$ be the value of the approximation at index r for fitting k optimal polynomial functions of order P between x_1 and x_r, given by

$$\beta_k(r) = \sum_{l=0}^{P} a_l^k r^l \tag{8}$$

where $\left\{a_l^k\right\}_{l=0}^{P}$ are the optimal polynomial coefficients of the k^{th} polynomial function for fitting k optimum polynomial functions of order P between x_1 and x_r ($kP+1 \leq r \leq N$).

Note that $D_1(r),\ (P+1 \leq r \leq N)$ can be obtained by minimizing a cost function similar to eqn. (3) without any constraint and setting $M_1 = 1$ and $M_2 = r$. Therefore, eqn. (6) can be used with proper modification[3] in this context. $D_k(r)$ is computed in a recursive manner and $\xi_k(r)$ and $\beta_k(r)$ are stored in each recursion of dynamic programming as described below.

Dynamic programming algorithm

1. Initialization:

Compute $D_1(r)$ and $\beta_1(r)$, $r = P+1, ..., N$. For each r, use eqn. (6) without any constraint and obtain $\{a_l^1\}_{l=0}^{P}$, which will be used in eqn. (8) to obtain $\beta_1(r)$. Also $\xi_1(r) = 1,\ r = P+1, ..., N$.

2. Iteration:

For $2 \leq k \leq K$ and $kP+1 \leq r \leq N$ compute the following:

$$\begin{aligned} D_k(r) &= \min_{1 \leq s \leq r-P} \{D_{k-1}(s) + \delta(s, r, \beta_{k-1}(s))\} \\ \xi_k(r) &= \arg\min_{1 \leq s \leq r-P} \{D_{k-1}(s) + \delta(s, r, \beta_{k-1}(s))\} \end{aligned} \tag{9}$$

where $\delta(s, r, \beta_{k-1}(s))$ is computed using eqn. (6) and (7) and $\beta_k(r)$ is also computed using eqn.

and x_r to fit a polynomial of order P. For $r < kP+1$, $D_k(r)$ is set to ∞.

[3]Without the constraint $\sum_{p=0}^{P} \alpha_p M_1^p = \beta$, matrix A in eqn. (6) is modified to the top left $(P+1) \times (P+1)$ submatrix of A given in eqn. (5). $\underline{\theta}$ and $\underline{b}$ are modified by taking first $P+1$ elements of those in eqn. (5).

(6) and (8). The maximum range of s in $\delta(s, r, \beta_{k-1}(s))$ can be $r - P$ because the number of data points between x_{r-P} and x_r is $P+1$, which is the minimum number of data points required to fit a polynomial of order P.

3. Termination and Backtracking:

After $D_k(r)$ and $\xi_k(r)$ are computed, we backtrack to obtain optimal piecewise segment boundaries from $\xi_k(r)$. $D_K(N)$ is the SSE of the approximation error for fitting K optimal polynomial functions of order P between x_1 and x_N. So $\xi_K(N)$ is the starting point of the K^{th} piecewise polynomial. Thus, $\eta_{K-1} = \xi_K(N)$. It means the $(K-1)^{\text{th}}$ piecewise polynomial functions should end at η_{K-1} and should start at $\xi_{K-1}(\eta_{K-1})$; so, we can recursively compute

$$\eta_k = \xi_{k+1}(\eta_{k+1}) \qquad k = K-2, K-3, ..., 2, 1 \tag{10}$$

Since in eqn. (1) we defined $\eta_0 = 1$ and $\eta_K = N$, (10) gives $\{\tilde{\eta}_k\}$, $k = 0, ..., K$ of the optimization problem of eqn. (2). The optimum values $\left\{\tilde{a}_l^k\right\}$ of eqn. (2) are now obtained as follows:

$\{\tilde{a}_l^1\}_{l=0}^P$ are obtained following the solution of eqn. (3) without any constraint and setting $M_1 = \tilde{\eta}_0 = 1$ and $M_2 = \tilde{\eta}_1$. $\left\{\tilde{a}_l^k\right\}_{l=0}^P$, $k = 2, ..., K$ are obtained using eqn. (6) with $M_1 = \tilde{\eta}_{k-1}$, $M_2 = \tilde{\eta}_k$ and $\beta = \beta_{k-1}(\eta_{k-1})$.

4 Experiment and Results

Six hundred sentences were randomly chosen from the TIMIT database [15] for our experiment, and the Robust Algorithm for Pitch Tracking (RAPT) based on the autocorrelation method [16] was used to extract pitch values over every 10 msec frame. Pitch values obtained by RAPT were used as references for objective evaluation of the proposed scheme. The DP based optimum pitch stylization was applied to the pitch contour of each voiced segment of all 600 utterances (a total of 4231 voiced segments). Mean square error between the reference pitch and the stylized pitch in each voiced segment was used as the metric for objective evaluation. As a baseline

method, the boundaries of the K piecewise segments were blindly placed uniformly over the duration of the voiced segments. We also obtained the stylized pitch using the directed graph (DG) approach [1] to compare against the proposed DP based stylization.

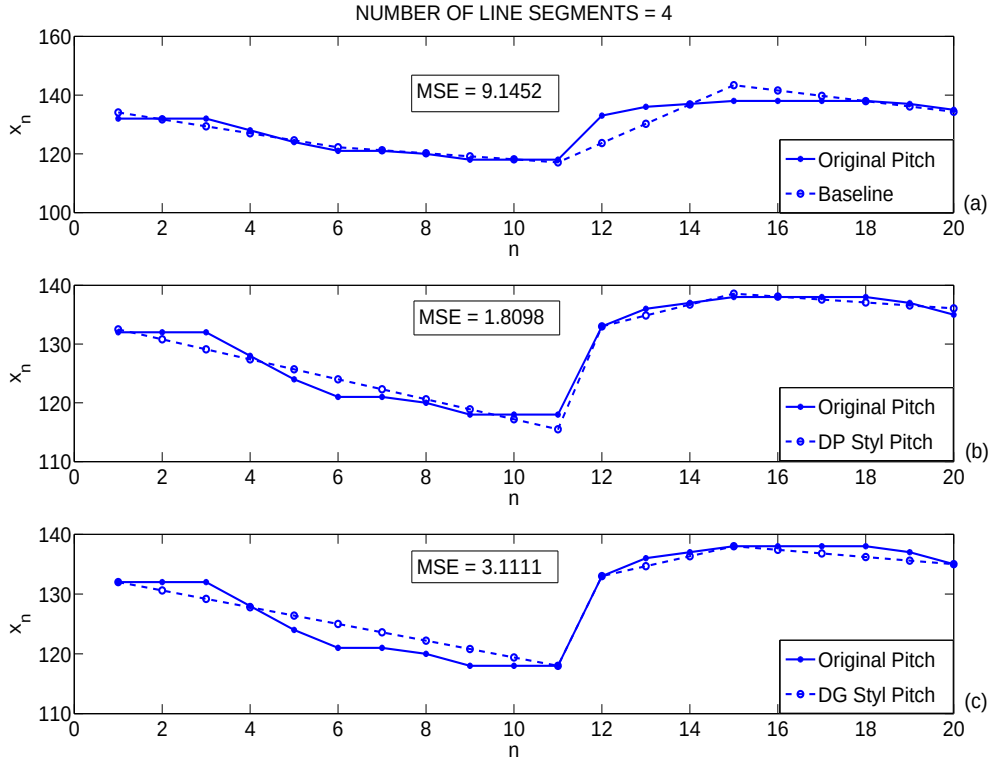

Figure 1: *Original and Stylized pitch contour (K=4 and P=1) using (a) baseline, (b) Dynamic Programming and (c) Directed Graph (DG) [1] approach.*

To obtain the stylized pitch values using piecewise polynomial functions, the number of piecewise segments K and the polynomial order P have to be provided. To determine the value of K for each voiced segment, we followed an approach similar to [5]. Wavelet decomposition of the pitch contour was performed using Daubechies wavelet (Db10), and the number of extrema in level 3 of the decomposition is used as $K-1$. Three different polynomial orders P were chosen - 1, 2, 3. For illustration, a sample pitch contour of a voiced segment and its stylization using baseline, DP, and DG approaches (K=4, P=1) are shown in Fig. 1 (a), (b) and (c) respectively. The MSEs are mentioned on the figures for comparison. It is clear that the stylization using DP

based approach achieves the best performance in terms of MSE.

For a comprehensive objective evaluation, average MSE over all the voiced segments of all the sentences are shown in Table. 1. It can be observed that DP based approach obtains the least MSE for all choices of P.

Stylization Schemes	Polynomial Order		
	P=1	P=2	P=3
Baseline	36.5145	13.2099	6.8429
DP	**5.1330**	**1.8967**	**0.7754**
DG	7.2706	2.1086	0.8167

Table 1: *Average MSE of pitch stylization using baseline, DP, and DG methods.*

For subjective evaluation, 6 sentences (3 male + 3 female) were randomly picked from the TIMIT database and their pitch contours were stylized using four different combinations of P and K in the DP based approach - *Combination 1:* $P1$=1 and $K1$ is obtained by Db10 decomposition, *Combination 2:* $P2$=2 and $K2$ is obtained by Db10 decomposition, *Combination 3:* $P3$=2 and $K3$=$\left\lceil \frac{(P1+1)K1}{P3+1} \right\rceil$ ($\lceil x \rceil$ is the smallest integer greater than x), and *Combination 4:* $P4$=3 and $K4$=$\left\lceil \frac{(P1+1)K1}{P4+1} \right\rceil$.

Total number of polynomial coefficients in Combinations 3 and 4 is the same as that of Combination1. Combinations 3 and 4 were intentionally chosen to check how the perception is affected by altering polynomial order but keeping the total number of parameters the same. All these stylized pitch contours were used to synthesize the utterances using PSOLA technique [17] and were compared against the original utterances through listening tests. In the listening test, the listeners were allowed to listen to an utterance as many times as they wanted and were asked to make a binary decision - whether the two utterances (original and synthesized) are perceived identical or not. Every pair of utterances was presented to the listener in a random order for different sentences. All the listeners were students 20-30 years old. Based on the decisions taken by 15 listeners, the percentages of listeners who thought two respective utterances are identical are shown in Table 2. It can be seen that most of the listeners found the original and synthesized utterances to be identical. The listening test results in combination 2 do not differ much from

Utterance No.	Combination of K and P			
	Comb.1	Comb.2	Comb.3	Comb.4
M1	93.33	93.33	86.67	100
M2	86.67	73.33	80.00	93.33
M3	93.33	93.33	80.00	86.67
F1	100	100	93.33	93.33
F2	100	80.00	100	93.33
F3	86.67	93.33	93.33	80.00

Table 2: *Result of Listening Test - percentage of people who perceived original and synthesized utterances are identical; M1, M2, M3 are three male speakers' utterances with durations 4.2, 2.7, 5.1 sec respectively; F1, F2, F3 are three female speakers' utterances with durations 3, 1.7, 2.7 sec respectively.*

those of combination 1. This is consistent with the observations made by Hart et al. [14]. It can also be noted that the listening test results for combinations 3 and 4 are not drastically different from those of combinations 1 and 2. The listening test results indicate listerners' tolerance to the quantization of the pitch contour representation by polynomial approximation. It should be noted that the result of the listening test using stylization obtained by DG approach turned out to be similar to that of DP approach, although DP achieves the minimum MSE.

5 Conclusions

The evaluation of the proposed DP based piecewise polynomial approximation of pitch contour shows that a stylized pitch contour which has minimum MSE for a given K and P also maintains perceptual closeness to the actual pitch contour. The DP based approximation technique makes it possible to change K and P and obtain different stylized versions of a pitch contour with the minimum MSE. This provides the flexibility to study, and potentially use, various parametric pitch stylizations within synthesis and speech modeling applications.

Acknowledgment

Work supported by NIH and NSF.

References

[1] R. Nygaard and D. Haugland, "Compressing ECG signals by piecewise polynomial approximation", *Proc. ICASSP*, vol. 3, May 1998, pp 1809-1812.

[2] Z. Han, S. Zhang, H. Zhang and B. Xu, "A vector statistical piecewise polynomial approximation algorithm for environment compensation in telephone LVCSR", *Proc. ICASSP*, vol. 2, April 2003, pp 117-120.

[3] S. Y.C. Catunda, O. R. Saavedra, J. V. FonsecaNeto and M. R.A. Morais, "Look-up table and breakpoints determination for piecewise linear approximation functions using evolutionary computation", *Proc. IMTC*, vol. 1, May 2003, pp 435-440.

[4] S. Ravuri and D. P.W. Ellis, "Stylization of pitch with syllable-based linear segments", *Proc. ICASSP*, April 2008, pp 3985-3988.

[5] D. Wang and S. Narayanan, "Piecewise linear stylization of pitch via wavelet analysis", *Proc. Eurospeech*, Lisbon, Portugal, October 2005, pp 3277-3280.

[6] D. Hirst and R. Espesser, "Automatic modelling of fundamental frequency using a quadratic spline function", *Travaux de l'Institut de phontique d'Aix*, vol. 15, 1993, pp 75-85.

[7] C. d'Alessandro and P. Mertens, "Automatic pitch contour stylization using a model of tonal perception", *Computer Speech and Language*, vol. 9, 1995, pp 257-288.

[8] P. Taylor, "The tilt intonation model", *Proc. ICSLP*, 1998, pp 1383-1386.

[9] A. Cantoni, "Optimal curve fitting with piecewise linear functions", *IEEE Trans. on Computers*, vol. C-20, no. 1, Jan 1971, pp 59-67.

[10] I. Tomek, "Two algorithms for piecewise-linear continuous approximation of functions of one variable", *IEEE Trans. on Computers*, vol. C-23, no. 4, April 1974, pp 445-448.

[11] M. Trajkovic and M. Hedley, "Recursive formulae for piecewise polynomial approximation of discrete functions", *Novi Sad J. Math.*, vol. 28, no. 2, 1998, pp 31-35.

[12] M. Obata, K. Wada, K. Toraichi, K. Mori and M. Ohira, "An approximation of data points by piecewise polynomial functions and their dual orthogonal functions", *Signal Processing*, vol. 80, 2000, pp 507-514.

[13] D. York, "Least-Square Fitting of a Straight Line", *Canad. J. Phys.*, vol. 44, 1966, pp 1079-1086.

[14] J. 't Hart, "F_0 stylization in speech: Straight lines versus parabolas", *J. Acoust. Soc. Am.*, vol. 90, issue 6, Dec 1991, pp 3363-3370.

[15] "DARPA-TIMIT", Acoustic-Phonetic Continuous Speech Corpus, NIST Speech Disc 1-1.1, 1990.

[16] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," *Speech coding and synthesis, W. B. Kleijn and K. K. Paliwal, Eds.: Elsevier Science*, 1995, pp 495-518.

[17] W. Roucos, and A. M. Wilgus, "High quality time-scale modification for speech," *Proc. IEEE ICASSP*, 1985, vol. 2, pp 493-496.