Educational Software for Optimal Order Determination of LP Speech Modeling in Telecommunication

ONDREJ RASO, MIROSLAV BALIK
Brno University of Technology
The Faculty of Electrical Engineering and Communication
Department of Telecommunication
Purkynova 118, 612 00 Brno,
CZECH REPUBLIC
Email: ondra.raso@phd.feec.vutbr.cz, balik@feec.vutbr.cz

Abstract: - This paper describes an algorithm from educational software family which was created in our department of telecommunication in the Brno University of technology. The software family is used for linear predictive modeling education purpose. This algorithm is used for optimal order length determination of LPC filter which was used for the modeling. This algorithm firstly finds an "optimal order" for each segment and then it selects "optimal order" for whole analyzed signal.

Key-Words: - Linear prediction model, education, speech signals, digital signal processing.

1. INTRODUCTION

Linear prediction (LP) modeling is used in wide area of applications and is often used technique for modeling speech signal in telecommunications. This type of modeling is used in the most modern speech codecs. These speech codecs are based on vocal tract modeling via linear predictive filter (LPC), so called LPC filter, whose input signal is either white noise or train of impulses with pitch period T [1], [2]. The most famous example of the use LPC for speech coding is Code-Excited Linear Prediction (CELP) coder which is in some variants used in GSM [3].

Software education tools were created for education purpose in our department of telecommunications. These education tools help students understand the deeper issue of speech signal linear predictive modeling. One of these tools is software for optimal order determination of LPC filter. An algorithm of this software will be described in the article. The algorithm is derived from standard theory of linear prediction and statistical signal processing. The main of contribution lies in implementation and experimental evaluation of this algorithm.

The correct LP order determination is often overlooked during a teaching, although it is very important for overall speech quality and minimal bit rate. Too high order increase codec bit rate and too low order on the other hand decrease speech quality [4], [5], [6]. Hence, it is important to find an optimal order for each group of analyzed signals.

2. LINEAR PREDICTION MODEL

A linear predictor forecasts the amplitude of the signal $x(n)$ at the discrete time n. It is using a linear combination of P previous samples

$$\hat{x}(n) = \sum_{j=1}^{P} a_j x(n-j)\,. \qquad (1)$$

where $\hat{x}(n)$ is the prediction of the signal $x(n)$ and a_j are coefficients of predictor of order P [1].

The linear prediction model is based on an idea, that the signal $x(n)$ is produced by all-pole filter $H(z)$ (2) with excitation signal $u(n)$ in the input, see Fig. 1.

$$H(z) = \frac{1}{1+\sum_{j=1}^{P} a_j z^{-j}} \qquad (2)$$

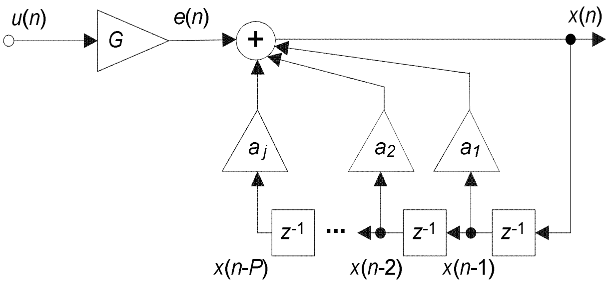

Fig. 1. Linear prediction model [1] .

The prediction error $e(n)$ is defined as

$$e(n) = x(n) - \sum_{j=1}^{P} a_j x(n-j) \quad (3)$$

The prediction error $e(n)$ is the difference between the signal $x(n)$ and its prediction $\hat{x}(n)$. From the previous equation (3), the signal $x(n)$ which is generated by the linear predictor can be synthesized as

$$x(n) = \sum_{j=1}^{P} a_j x(n-j) + e(n) \quad (4)$$

The coefficients a_j are obtained by minimizing a mean square error criterion function, i.e. minimizing a formula $\frac{\partial}{\partial \mathbf{a}} \mathrm{E}\{e^2(m)\}$, where $\mathbf{a}^{\mathrm{T}} = [a_1, a_2, ..., a_P,]$ and E{} is averaging [1] . The solution could be obtained for example by the popular Levinson-Durbin recursive algorithm.

A. Segmental Optimal Order

The segmentation is necessary in such a case for processing the long-term audio signal. Segmentation divides this signal into short stationary segments. These segments will be noticed by the index i, $i = 1,2, ..., I-1$, where I is total number of the segments. The actual segment of the signal $x(n)$ will be noticed as $x_i(n)$ and the previous segment will be $x_{i-1}(n)$ and so on.

It is obvious, that the prediction error energy E, defined as

$$E = \sum_{n=0}^{N-1} e^2(n) \quad (5)$$

where N is length of segment, decrease with increasing order P of LP model. The illustration of a typical prediction error energy E is shown in Fig. 2.

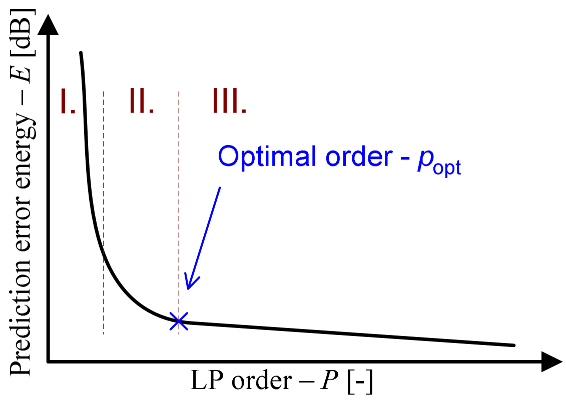

Fig. 2. Illustration of a typical prediction error energy E.

It can be seen from this figure, that the dependence has three different areas. In the first area, the prediction error decrease is the most significant. This area is followed by area of transition. In the last area the prediction error energy decrease is very slow. The optimal order p_{opt} will be chosen as an order between second and the last one area, i.e. higher order then optimal order p_{opt} will not decrease the prediction error too much.

Thus such smooth dependence from Fig. 2 can be obtained only as average from more than one realizations of prediction error $e(n)$, i.e. more measurements of signal $x(n)$ must be done. It is very problematic to get more then one realization of $x(n)$ in practice. In the speech signal, the same speaker would have had to say the same sentence or word with the same loudness, same emotion etc. The situation is even more complicated for the segmentation case. In the rest of the paper, prediction error $e(n)$, prediction error energy E and optimal order p_{opt} will be calculated for each segment and they will be noticed as $e_i(n)$, $E(i)$ and $p_{\mathrm{opt}}(i)$.

Fig. 3 shows prediction error energy $E(i)$ measured in real conditions, trumpet playing record with segment length $N = 512$ samples.

ISBN: 978-1-61804-109-8

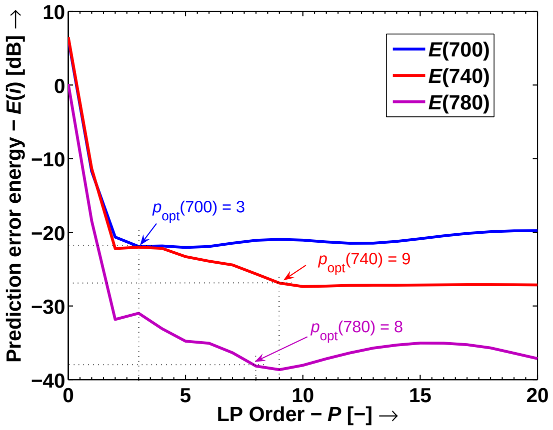

Fig. 3. Prediction error energy $E(i)$ for three different segment, the arrows mark optimal order $p_{opt}(i)$ for each segment.

This figure shows three curves for three different segments, i.e. i = 700, 740, 780. The arrows mark the chosen optimal order $p_{opt}(i)$ and its value for each segment. The optimal order was chosen by a following method. Firstly, the curves running of prediction error energy $E(i)$ are "smoothed" by the median filter of length 3, see Fig. 4.

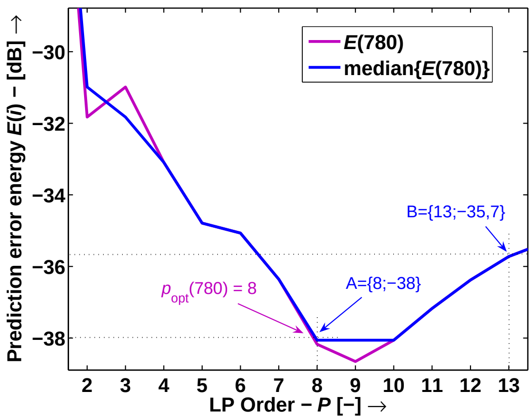

Fig. 4. Detail of $E(780)$ from Fig.3 (pink curve) and smoothed curve by median filter of length 3 (blue curve).

The smoothed curve running is than observed. Energy of prediction error with higher order should be lower than with lower order, if not or the decreased is not significant, the order increasing is useless. This situation is illustrated in Fig. 4. The point B has higher order than A, but the point B has lower energy even with the higher order. Hence, the order of point A was chosen as optimal, the pink arrow from Fig. 4.

The flowchart in Fig. 5 exactly describes the algorithm for the optimal order $p_{opt}(i)$ choosing. The index-i is missing for a better overview in this flowchart.

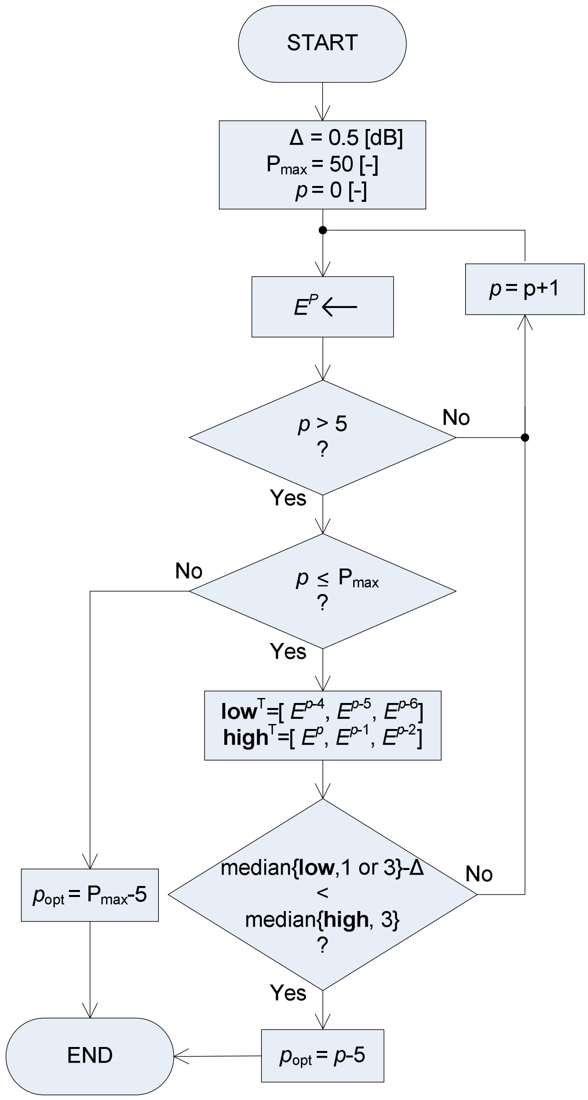

Fig. 5. Algorithm of optimal order $p_{opt}(i)$ choosing method.

The algorithm firstly calculates the first five prediction errors energy for the first five orders. It is marking by symbols $E^{p} \leftarrow$, i.e. the prediction error energy E of p order. Then the algorithm compares two vectors of prediction error energy, see $\mathbf{low}^{T}$ and $\mathbf{high}^{T}$ vectors in Fig.5 . If the prediction error energy is higher in the vector $\mathbf{high}^{T}$ than in the vector $\mathbf{low}^{T}$ then the algorithm chooses as the optimal order the current order p minus 5, because the difference between these two vectors is 5 orders. The median filter of length three is applied to the each vector before the comparison.

B. Global Optimal Order

In the previous subsection, the method for the optimal order $p_{opt}(i)$ choosing was described. This

ISBN: 978-1-61804-109-8

order is optimal only for one segment. Hence, optimal order selection has to be chosen not for only one segment but for whole long-term audio signal. In Fig.6, optimal orders $p_{opt}(i)$ illustration of first 150 segments are shown. It can be seen, that the optimal orders $p_{opt}(i)$ of these segments are very variable, they are from 2 up to 21 order. A quantiles were used for optimal order of whole long-term audio signal. In Fig. 6, four quantiles are shown:

- $p_{0.5}$, 50% quantile (or median), blue curve.
- $p_{0.75}$, 75% quantile, green curve.
- $p_{0.95}$, 95% quantile, red curve.
- $p_{0.99}$, 99% quantile, pink curve.

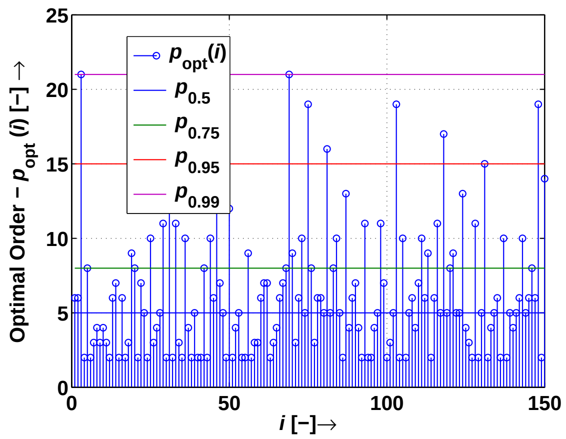

Fig. 6. Illustration of optimal orders $p_{opt}(i)$ of first 150 segments, the audio signal is short record of trumpet playing.

An illustration of sorted LP orders is shown in Fig. 7. 95% quantile p0.95 was experimentally chosen as global optimal order, i.e. optimal order for whole analyzed long-term audio signal.

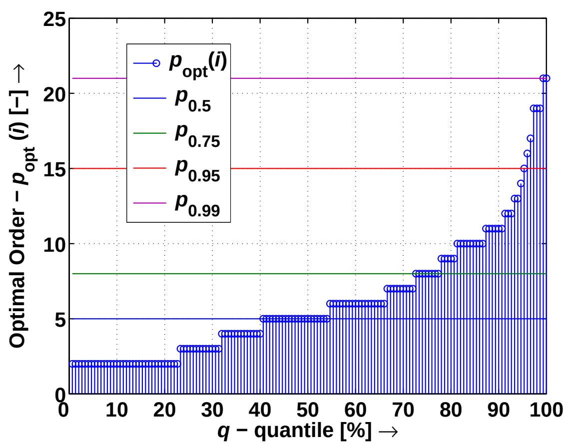

Fig. 7. Illustration of sorted LP orders.

3. Software Implementation

The algorithm which was described above sections was implemented in Matlab as a run-time script.

4. Conclusion

Software implantation of the algorithm for optimal order determination of linear predictive modeling of speech signal was described in this article. This software is part of group of software for education purpose which was created in our department of telecommunication.

This algorithm is based on measuring prediction error energy in the analyzed signal. This algorithm firstly finds an “optimal order” for each segment and then it selects “optimal order” for whole analyzed signal.

5. Acknowledgements

This work has been supported by the grant OPVK no CZ.1.07/2.2.00/28.0062, joint activities of BUT and TUO while creating the content of accredited technical courses in ICT.

References:

[1] VASEGHI, S. V.: *Advanced Digital Signal Processing and Noise Reduction*. 2nd edition. New York : Wiley, 2000. 473 p. ISBN 0-471-62692-9.

[2] JAN, J.; *Číslicová filtrace, analýza a restaurace signálů* (in czech). 2nd edition. Brno, Vuitium, 2002. 427 p. ISBN: 80-214-1558-4.

[3] HÄNSLER, E., SCHMIDT G., *Topics in Acoustic Echo and Noise Control: Selected Methods for the Cancellation of Acoustic Echoes, the Reduction of Background Noise, and Speech Processing*. Berlin-Heidelberg-New York: Springer, 2006. 642 p. ISBN 10 3-540-33212-x.

[4] HÄNSLER, E. S.; SCHMIDT, G.:. *Speech and Audio Processing in Adverse Environments*. Berlin-Heidelberg-New York: Springer, 2008. 735 p. ISBN 978-3-540-70601-4. FASTL, H.,

[5] ZWICKER E., *Psychoacoustic: Facts and Models. 3rd ed. Berlin-Heidelberg-New York*: Springer, 2006. 462 p. ISBN 978-3-540-23159-2.

[6] BALÍK, M., RASO, O., Sound Artifacts Caused by the Thresholding Method. *In Proceedings of the IASTED International Conference on Modeling, Simulation, and Identification*. ACTA Press, Canada: ACTA Press, 2009. s. 1-7. ISBN: 978-0-88986-810- 6.