

# Linear Prediction

## 1 Questions

1. At first, we want to discuss the source-filter model. Have a look on slides 4 and 5. Answer the following questions for the human body and the model, respectively!
  - a) Where is the filter part located?
  - b) Where is the source part located in the case of voiced speech?
  - c) Where is the source part located in the case of unvoiced speech?
2. Which cost function is minimized by the linear prediction?
3. What property of a signal is exploited by a predictor?
4. In the lecture, the principle of linear prediction is applied in form of a 'short term' predictor. Literature and various applications also define a 'long term' predictor. What is the difference between these two predictors and why is a distinction necessary/advantageous?
5. What problems (may) occur when straightforwardly solving the set of normal equations?
6. The Levinson-Durbin algorithm presents an alternative to straightforwardly solving the equation system. What exactly is advantageous? What is increased in each iteration?
7. What are the "break" conditions for the Levinson-Durbin algorithm?
8. What are the properties of the resulting prediction-error filter and its inverse filter?
9. Assume you want to use linear prediction for speech transmission. What do you have to transmit from sender to receiver in order to perfectly reconstruct the speech signal  $s(n)$ ?
10. In what ways does the application of linear prediction for speech transmission have advantages over straightforward quantization or pulse-code modulation?

## 2 Answer

1. a) Human body: In the pharynx, nasal and mouth cavity (orange part slide 4)  
Model: Orange part slide 5
- b) Human body: In the lungs and vocal cords (green part slide 4)  
Model: Impulse generator (green part slide 5)
- c) Human body: Depending on the kind of unvoiced sound (/f/: lips, /sch/: tongue, /h/: vocal cords) (green OR orange part slide 4)  
Model: Noise generator (green part slide 5)
2. The expectation of the squared error (between the desired signal and the estimated desired signal) is minimized. (See slide 9)
3. The correlation within the signal.
4. Due to the filter-part in the filter-source model of speech production, the autocorrelation function of a speech signal yields large values in a 'short' range around zero lag. This correlation is removed by the predictor filter described in the lecture. For voiced speech segments, however, the periodicity of the excitation-part in the source-filter model also yields large autocorrelation values at lags corresponding to multiples of the pitch period. A predictor aiming at only removing the correlation induced by the pitch period is referred to as 'long term' predictor. Without this distinction, a prediction filter with a significantly large number of filter taps would need to be computed to counteract both 'types' of correlation.
5. High complexity + numerical problems.
6. Due to the iteration no matrix inversion is required. The filter order is increased in each iteration.
7. Reaching the desired (or maximum) filter order is the first break condition. One should also abort if numerical problems occur (complete reflection, increase in error power). It is also possible to stop the iteration before the maximum filter order is reached if a "pseudo-convergence" is achieved (relative decrease of error power below a threshold).
8. An FIR-Filter is computed with all its zeros inside the unit circle, signals can pass the filter with a minimum delay (minimum phase property), the inverse prediction filter (IIR) is stable since all zeros become poles and, thus, are located inside the unit circle.
9. You have to transmit the error signal  $e(n)$  and the filter coefficients  $h_i$ .
10. Due to linear prediction, the dynamic range of the signal to be transmitted is significantly reduced (see prediction gain). Therefore, fewer bits are required for transmission of the signal. Since the predictor coefficients are only transmitted once per frame, this, overall, yields a bitrate reduction. If a degradation of speech quality is acceptable, transmission may also be purely parametric (no error signal, instead information on voiced/unvoiced etc.).