3 Speech Coding: Psychoacoustics and Audio Coding

3.1 Overview

In this exercise, we carry out experiments on

- Linear prediction: forward vs. backward
- Psychoacoustics: the foundation for audio coding (e.g. *MPEG audio*).

3.2 Speech Coding

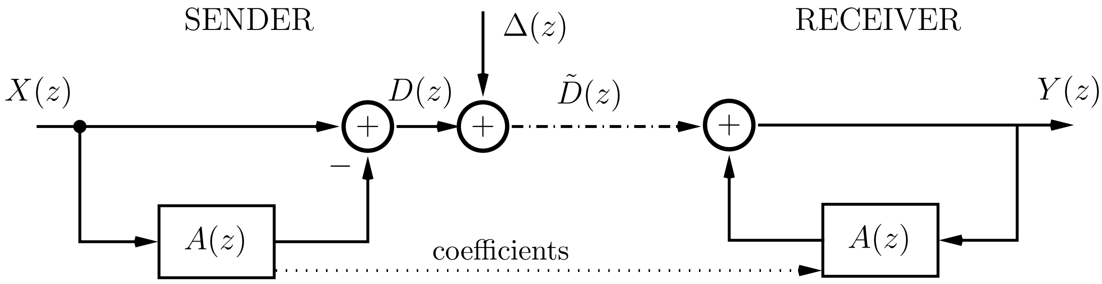

Fig. 4: Forward prediction

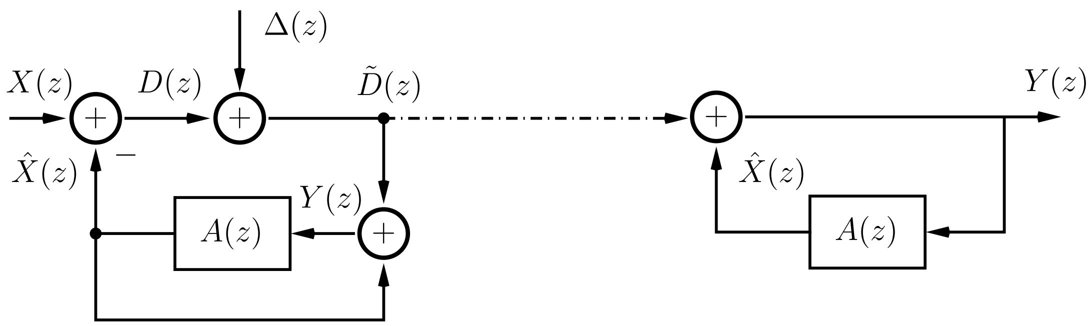

Fig. 5: Backward prediction

$$
\begin{aligned}
Y_{\text{fw}}(z) &= [X(z)\,(1 - A(z)) + \Delta(z)]\,\frac{1}{1 - A(z)} && (10)\\
&= X(z) + \frac{\Delta(z)}{1 - A(z)} && (11)\\
Y_{\text{bw}}(z) &= X(z) - \hat{X}(z) + \Delta(z) + \hat{X}(z) && (12)\\
&= X(z) + \Delta(z) && (13)
\end{aligned}
$$

3.2.1 Preparation

Before starting with the experiments, answer the following questions: T

a, There are generally two options for linear prediction:

- forward prediction (*open loop*, *Fig.* 4), and
- backward prediction (*closed loop*, *Fig.* 5).

What are the advantages and disadvantages of both methods?

▷ ____

▷ ____

b, How are both methods related?

▷ ____

c, Which method exhibits a higher signal-to-quantization error power ratio at the decoder output?

▷ ____

3.2.2 Forward Prediction, Noise Shaping

The following lines (see file `lpc_forw.m`) show a MATLAB code of a simple forward linear predictor (*Fig* 4). It works with overlapping blocks of length $160 + 80 = 240$, where 80 samples are overlapping (typical values, e.g., AMR coder for UMTS).

a, Compare the source signal `s`, signal `err2`, and output signal `y`. Which part of the linear model of speech production (excitation, signal shaping in the vocal tract, ...) is represented by the signal `err2`? M

▷ ____

b, The above program calls the `lpc` command.

- Which method is used, autocorrelation or covariance method?

  ▷ ____

- Is the underlying autocorrelation estimate biased or unbiased?

  ▷ ____

c, Derive (theoretically) the transfer function $G(z)$ of the decoder part in the z domain.

▷ ____

```
% ===========================================================
% Basic Linear Predictive Coding (forward)
% ===========================================================

% -------------
% input signal
% -------------
s = audioread('male.wav'); % sampling rate 8kHz

% -----------
% LP encoder
% -----------

L1 = 160; % block length
L2 = 80;  % overlap (previous block)
N  = 10;  % order of predictor

A = [];     % matrix for predictor  coefficients

err  = zeros(size(s,1),1);
y    = zeros(size(s,1),1);

no_iterations = floor(size(s,1)/L1)-1;    % division due to block processing

for m = 2:no_iterations
  % windowed frame (including samples from past frame)
  s_p = hamming(L1+L2) .* s( L1.*(m-1)-L2 : L1.*m-1 );

  a = real(lpc(s_p,N));    % N+1 by 1 coefficient vector
  A = [A; a];

  % convolution
  for k = 0:L1-1
    err(L1.*(m-1)+k) = a * s(L1.*(m-1)+k:-1:L1.*(m-1)-N+k);
  end
end

% --------------
% Quantization
% --------------

err2 = quant(err,0.001);

% -----------
% LP decoder
% -----------
% Equation y(k)=err2(k)+a*y(k)

for m = 2:no_iterations
  % convolution, synthesis  filter
  for k = 0:L1-1
    y(L1.*(m-1)+k) = err2(L1.*(m-1)+k)...
              - A(m-1,2:N+1) * y(L1.*(m-1)+k-1:-1:L1.*(m-1)+k-N);
  end
end
```

3.3 Psychoacoustics

While speech coding exploits properties of the human speech production, *audio coding* (suitable for more general signals) exploits properties of the human ear (*psychoacoustics*). We refer to the lecture notes, p. 33ff.

There are two *masking* phenomena:

- Masking in the frequency domain
- Masking in the time domain

The following simple experiments can be carried out with MATLAB.

3.3.1 Masking in the Frequency Domain

The *hearing area* is commonly described as a nonlinear function of frequency and sound pressure level (SPL) or sound intensity I.

Preparation

T

a, Sketch such a diagram containing the *threshold in quiet*. At what frequency is the highest sensitivity?

b, Let us assume now the presence of a sinusoidal signal of 1kHz. What does the corresponding masking threshold look like?

Experiments

M

a, Open a new MATLAB file. Create a sinusoidal signal (the masker, 1kHz, sampling frequency 8kHz, duration 1s) and listen to it using `sound`.

b, Add another sinusoidal signal with linearly increasing frequency and constant amplitude (much lower than that of the masker) and listen to the resulting signal. You may use the function `chirp` to create the sinusoidal signal.

c, Add a sinusoidal signal with varying amplitude and constant frequency (e.g., 1.1kHz) to the masker and listen to the resulting signal.

d, Listen to the sum of a sinusoidal with frequency 1kHz and a sinusoidal with frequency 1.4kHz. What effect do you observe?

e, Listen to a sinusoidal signal with frequency 500Hz at the left ear and a signal with frequency 501Hz at the right ear. What effect do you observe?

3.3.2 Masking in the Time Domain

There is pre-masking when the masker sets in, simultaneous masking and post-masking after the maker is switched off.

Preparation

T

a, How can pre-masking be explained?

▷ ____________________

Experiments

M

a, Open a new MATLAB file. Create a stationary white noise sequence ($1s$). Append zeros on both ends ($1s$ each).

b, Add short impulses with varying distance to the start of the noise sequence and listen to the signals.

c, Add short impulses with varying distance to the end of the noise sequence and listen to the signals.