Applications of Linear Prediction

Slides for this lecture are based on those created by Katariina Mahkonen for TUT course "Puheenkäsittelyn menetelmät" in Spring 2013.
Recap about linear prediction

• Linear prediction is one of the most important speech processing tools.

• From the speech processing viewpoint, the most important characteristic of LP is its ability to model the vocal tract.

• The idea is to predict the next sample of a speech signal as a linear combination of preceding samples (linear filter).

• Previously-discussed lattice-structured model for the vocal tract is an all-pole filter → linear prediction is a good method for estimating the parameters of the filter.
Recap about linear prediction

\[ S(z) = G(z)H(z) \]
The vocal tract filter \( H(z) \) can be modeled as all-pole filter

\[ H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \ldots + a_p z^{-p}}. \]

* The output of an all-pole system can be predicted perfectly if the latest input and the previous outputs are known. Since the glottis signal (input) is generally not known, we predicting approximately, based on the output (recorded speech) only

\[ \hat{y}(n) = - \sum_{k=1}^{p} a(k) y(n-k) \]

so that the sum of squared error \( E \) is minimized:

\[ E = \sum_{n=1}^{N} (y(n) - \hat{y}(n))^2 \]

\[ \hat{a}_{OPT} = \arg \min_{\hat{a}} \sum_{n=1}^{N} (y(n) - \hat{y}(n))^2 \]

* Vector \( \hat{a}_{OPT} \) is called linear prediction coefficients
Recap: linear prediction and Yule-Walker equation

Optimal filter parameters (LP coefficient) \( a(1), a(2), \ldots, a(p) \) are found by setting the partial derivatives of \( E \) with respect to each parameter \( a(n) \) to zero.

Represented with the autocorrelation function \( r(n) \), the zeros of the partial derivatives can be written in matrix form as the Yule-Walker equation:

\[
\begin{bmatrix}
    r(0) & r(1) & r(2) & \cdots & r(p-1) \\
    r(1) & r(0) & r(1) & \cdots & r(p-2) \\
    r(2) & r(1) & r(0) & \cdots & r(p-3) \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    r(p-1) & r(p-2) & r(p-3) & \cdots & r(0)
\end{bmatrix}
\begin{bmatrix}
    a(1) \\
    a(2) \\
    a(3) \\
    \vdots \\
    a(p)
\end{bmatrix}
=
\begin{bmatrix}
    r(1) \\
    r(2) \\
    r(3) \\
    \vdots \\
    r(p)
\end{bmatrix}
\]

- Levinson-Durbin recursion is a fast algorithm for solving the Yule-Walker equations.
Using LP coefficients to estimate formants

- Formant is a resonance of the vocal tract (or, pole of the vocal tract transfer function) that is visible in the speech spectrum
Formant estimation: factorization of the LP polynomial

- A straightforward way to estimate formants is to factorize the LP polynomial

\[ A(z) = 1 + a_1 z^{-1} + ... + a_p z^{-p} \]

into factors

\[ A(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1})...(1 - z_p z^{-1}), \]

where \( z_1, z_2, ..., z_p \) are the roots of the LP polynomial.
Magnitude response of a pole pair: formant frequency

- Pair of poles $re^{\pm j\theta}$ has transfer function:

\[
\frac{1}{(1 - re^{j\theta}z^{-1})(1 - re^{-j\theta}z^{-1})} = \frac{1}{1 - r(e^{j\theta} + e^{-j\theta})z^{-1} + r^2e^{j\theta}e^{-j\theta}z^{-2}} = \frac{1}{1 - 2r\cos(\theta)z^{-1} + r^2z^{-2}}
\]

- So the coefficients of the transfer function are:

\[
\begin{align*}
a_0 &= 1 \\
a_1 &= -2r\cos(\theta) \\
a_2 &= r^2
\end{align*}
\]

- On the unit circle $z = e^{j\omega}$ in the complex plane, the transfer function can be written as:

\[
\frac{1}{(1 - re^{j\theta}z^{-1})(1 - re^{-j\theta}z^{-1})} = \frac{1}{(1 - re^{j\theta}e^{-j\omega})(1 - re^{-j\theta}e^{-j\omega})} = \frac{1}{(1 - re^{j(\theta-\omega)})(1 - re^{-j(\theta+\omega)})}
\]

- The magnitude response (absolute value of the transfer function) gets its maximum value when $(1 - re^{j(\theta \pm \omega)})$ gets its minimum value, and that happens when $e^{j(\theta \pm \omega)} = 1$ from which we get $\omega = \pm \theta$ corresponding to frequency $\pm \frac{\theta}{2\pi} F_s$. 


Magnitude response of a pole pair: formant bandwidth

- Format bandwidth expresses how wide a formant is: if a formant is steep, its bandwidth is small, and vice versa

- Bandwidth is defined as the frequency range between the points where the magnitude response has dropped 3dB from its maximum value

- The bandwidths of a pole pair $re^{\pm j\theta}$ depends on the distance of the pole from the origin
  \[ BW_\omega = -2 \ln r \quad BW_f = -2 \ln r \frac{F_s}{2\pi} \]
Magnitude response of a pole pair: formant frequency and bandwidth

- Pole pair  \( z = 0.95e^{\pm i \cdot 0.1 \cdot 2\pi} \)

\[
f = \frac{0.1 \cdot 2\pi}{2\pi} 16000 \text{Hz} = 1600 \text{Hz}

\[
BW_f = -2 \ln(0.95) \cdot \frac{16000}{2\pi} = 260 \text{Hz}
\]
Example

Spectrogram of the words "Suomen laki" and formants (crosses) estimated from 8th order LP model [Koppinen2006].
Formants can be removed and the glottal excitation $G(z)$ solved by filtering a speech frame with an inverse filter $A(z)$ that involves the parameters from LP analysis:

\[
S(z) \xrightarrow{A(z)} G(z)
\]

- Fundamental frequency can be more reliably computed from the glottal excitation signal $G(z)$ than from speech frame $S(z)$ directly ⇒ Influence of formants is reduced

- Often autocorrelation function of $G(z)$ is computed within the frame and the maximum of the autocorrelation function is sought in the feasible range of fundamental frequencies

\[
S(z) = G(z)H(z) = \frac{G(z)}{A(z)}
\]

\[
g(z) = S(z)A(z)
\]
Glottal excitation signal using inverse filtering

phone /a/

- Original signal
- Glottal signal

Time (ms)