Formant Estimation from Noisy Voiced Speech

Ashok Krishnamurthy
Jian Li
Randy Moses
Department of Electrical Engr.
The Ohio State University
Closed phase speech signal model

- Speech signal in the closed phase is the free response of a linear system.
- Let $s(n), \ n = 0, \ldots, N - 1$ be speech signal in closed phase. Then
  \[ s(n) = \sum_{i=1}^{K} A_i e^{-\alpha_i n} \cos(\omega_i n + \phi_i) + e(n). \]
- The formant frequencies are
  \[ \frac{\omega_i F_s}{2\pi} \text{Hz}, i = 1, \ldots, K. \]
  $F_s$ is the sampling frequency in Hz.
- The formant bandwidths are
  \[ \frac{\alpha_i F_s}{\pi} \text{Hz}, i = 1, \ldots, K. \]
- The energy of each formant mode in the closed phase is
  \[ \frac{(1 - |z_i|^{2(\frac{N-1}{2})})|c_i|^2}{1 - |z_i|^2}, \ i = 1, \ldots, K. \]
  where $z_i = e^{-\alpha_i + j\omega_i}$ and $c_i = A_i e^{j\phi_i}$. 

2
Solution Procedure


- First step:
  Solve backward linear prediction equations:

\[
\begin{bmatrix}
  s(1) & s(2) & \cdots & s(L) \\
  s(2) & s(3) & \cdots & s(L+1) \\
  \vdots & \vdots & \ddots & \vdots \\
  s(N-L) & s(N-L+1) & \cdots & s(N-1)
\end{bmatrix}
\begin{bmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_L
\end{bmatrix}
= \begin{bmatrix}
  s(0) \\
  s(1) \\
  \vdots \\
  s(N-L-1)
\end{bmatrix}
\]
Solution Procedure (contd.)

- Second step:
  Find roots of $B(z) = z^L + b_1 z^{L-1} + \cdots + b_L$. The formants are a subset of these roots, and this gives the formant frequencies and bandwidths.
  Use the roots to obtain the amplitudes, phases and energies of the formants.
- Retain $K$ highest energy modes as the formants.
Modifications to Parthasarathy & Tufts procedure

- Use the electroglottograph signal to locate the closed glottal phase.

- Use total least squares in solving \( Yb = -y \).

- Use data from multiple consecutive closed phases (typically 3). \( Y_1b = -y_1 \), \( Y_2b = -y_2 \) and \( Y_3b = -y_3 \) are combined to yield

\[
\begin{bmatrix}
Y_1 \\
Y_2 \\
Y_3
\end{bmatrix} b = - \begin{bmatrix}
y_1 \\
y_2 \\
y_3
\end{bmatrix}.
\]
Example Results

- Normal, male speaker.
- Analyzed steady vowel portion from "bach".
- 16 bit A/D, sampling frequency 10 KHz.
- Simultaneously recorded speech and electroglottograph signals.
Formant contour, no noise

![Graph showing formant contours without noise.](image-url)
Energy of different formant modes

- speech signal
- first formant
- second formant
- third formant
- three highest energy formants
Comparison of original and synthesized speech

No noise.

Only 3 highest energy formants used in synthesis.
Formant contour, 20 dB SNR
Comparison of original and synthesized speech

20 dB SNR.

Only 3 highest energy formants used in synthesis.
Formant contour, 10 dB SNR

10 dB SNR: 3 closed phase analysis

Freq (Hz)

Time (ms)
Comparison of original and synthesized speech

10 dB SNR.

Only 3 highest energy formants used in synthesis.
Conclusions

• Very effective method of formant estimation for clean speech signal.

• Three formants appear to account for most (90%) of the energy in the closed phase for vowels.

• Performance is good with noisy speech signals upto 20 dB SNR.