Source filter modeling and spectral envelope estimation

summer 2006 lecture on analysis,
modeling and transformation of audio signals

Axel Röbel
Institute of communication science TU-Berlin
IRCAM Analysis/Synthesis Team

25th August 2006

Contents

1 Physical background

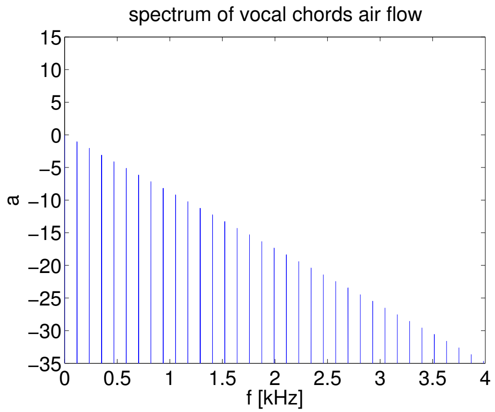
Source - Filter Model of Sound production

- separate pitch from timbre
- a physically founded sound production model that can be used manipulate many sound sources
- Source : harmonic excitation signal with flat (white) frequency spectrum - pitch information

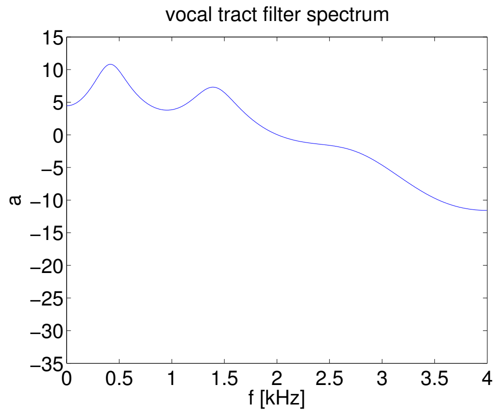
  physical correlate : vibrating lips (trumpet), vocal chords (speech), string (guitar, violin) reed (sax, clarinet).
- Filter : attenuates or amplifies energy

  physical correlate : vocal tract (speech), instrument body, resonator (musical instrument)

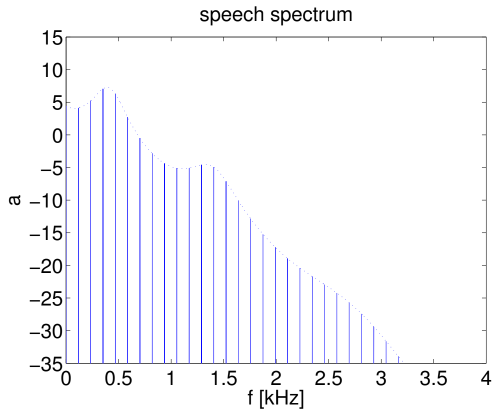
1.1 simplified physical model for voiced speech


spectrum of vocal chords air flow

vibrating vocal chords create periodic air pulses with lowpass character and desired fundamental frequency.


vocal tract filter spectrum

vocal tract filter creates formants

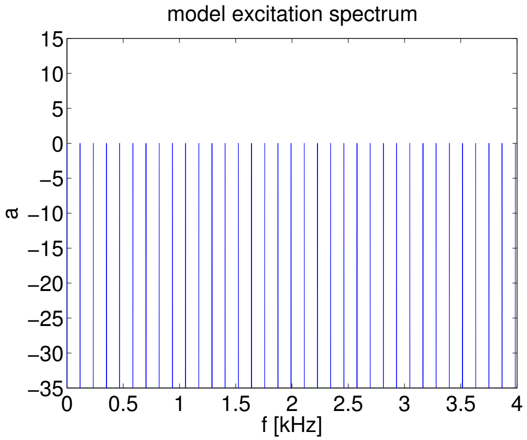

speech spectrum

spectrum of voiced speech spectrum

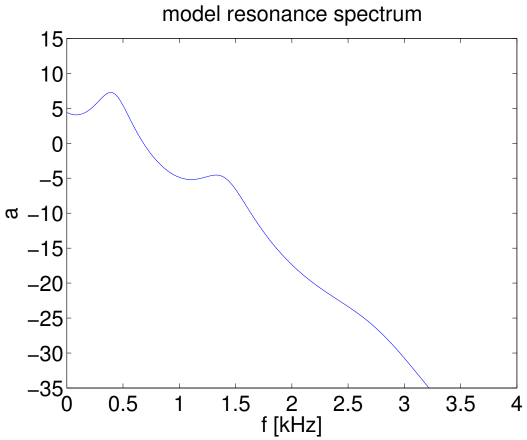
approximate model of pitch changes

- change the fundamental frequency of the air flow puls train,
- keep form of glottic air pulses (lowpass) and vocal tract formant structure,

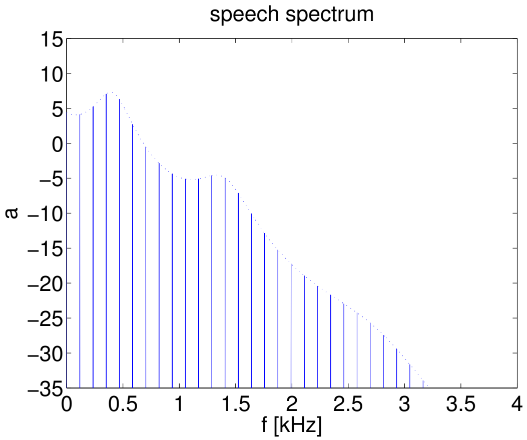
1.2 Source - filter model of voiced speech



model excitation spectrum, flat envelope with desired fundamental frequency



resonator filter incorporates glottic puls lowpass + vocal tract resonances



voiced speech spectrum

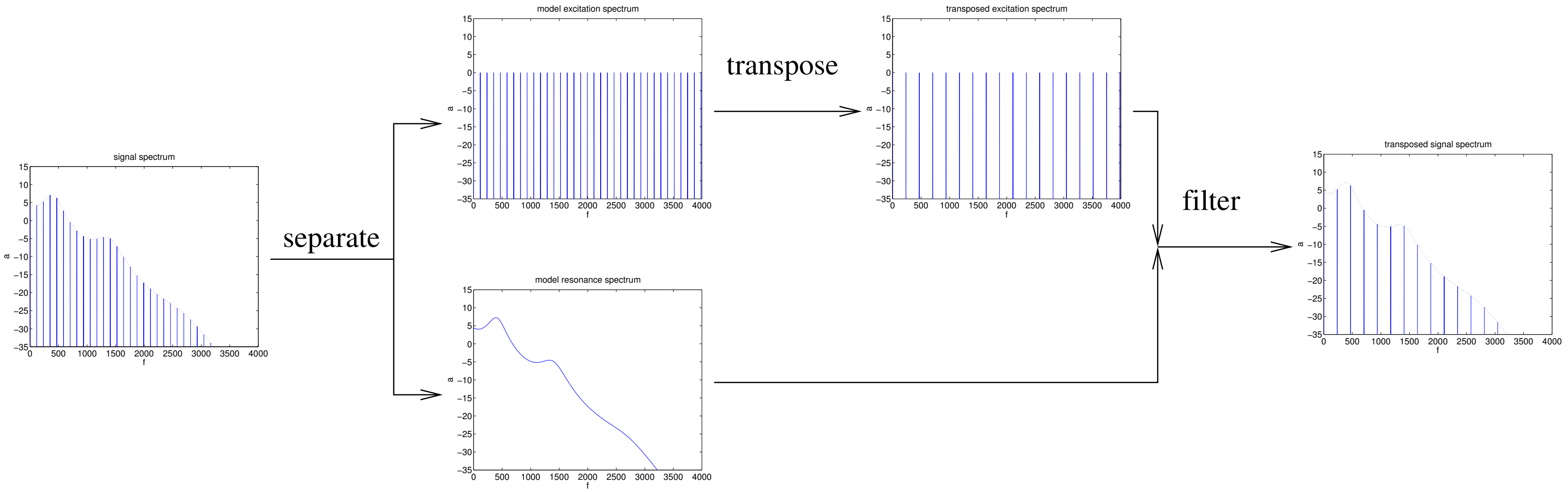
Reorganization with respect to physical model:

- concentrate pitch effects in excitation block (Source)
- combine spectral envelope effects (glottic puls lowpass + vocal tract resonances) in filter block (Filter)

1.3 sound transposition

Signal transposition strategy:

1. separate signal into **resonator filter** and **periodic and white excitation signal**.
2. transpose excitation to perform pitch shifting
3. re-apply **resonator filter** to reestablish original vocal tract resonances and the form of the glottic pulses.



2 Envelope estimation

There exist a number of methods for the estimation of the spectral envelope:

- Estimation of AR model parameters : linear prediction, discrete all pole modeling
- Estimation of cepstral parameters : discrete cepstrum, true envelope

We will discuss the implications of the different methods in the following

3 AR Model

The auto-regressive (AR) or linear predictive process explains a signal in terms of a white noise source and linear combinations of previous values of the signal

$$s(n) = (\sum_{k=1}^{P} a_k s(n-k)) + u(n) \tag{1}$$

- s is the observed signal
- u is the excitation signal
- P is the order of the model
- a_k are the model coefficients

the spectral relation between input u and output signal s can be obtained

$$\begin{aligned}
s(n) &= (\sum_{k=1}^{P} a_k s(n-k)) + u(n) && (2)\\
u(n) &= s(n) - \sum_{k=1}^{P} a_k s(n-k) && (3)\\
u(n) &= \sum_{k=0}^{P} a'(k) s(n-k) \quad \text{with} \quad a'(0) = 1, a'(k) = -a_k \forall k > 0 && (4)\\
u(n) &= s(n) * a(n) && (5)\\
U(w) &= S(w)A(w) = S(w) \sum_{k=0}^{P} b(k) e^{-jwk} && (6)\\
S(w) &= \frac{U(w)}{A(w)} && (7)
\end{aligned}$$

The transfer function of the linear prediction model is an all pole filter.

- for stable filter all poles are inside the unit circle.
- for all pole filters with $a_0 = 1$ there exist the following property [MG76]

$$\int_{-\pi}^{\pi} \log(|A(w)|^2) dw = 2 \int_{-\pi}^{\pi} \log(|A(w)|) dw = 0 \tag{8}$$

- this means that the area under the transfer function is always equally distributed above and below the 0dB line.

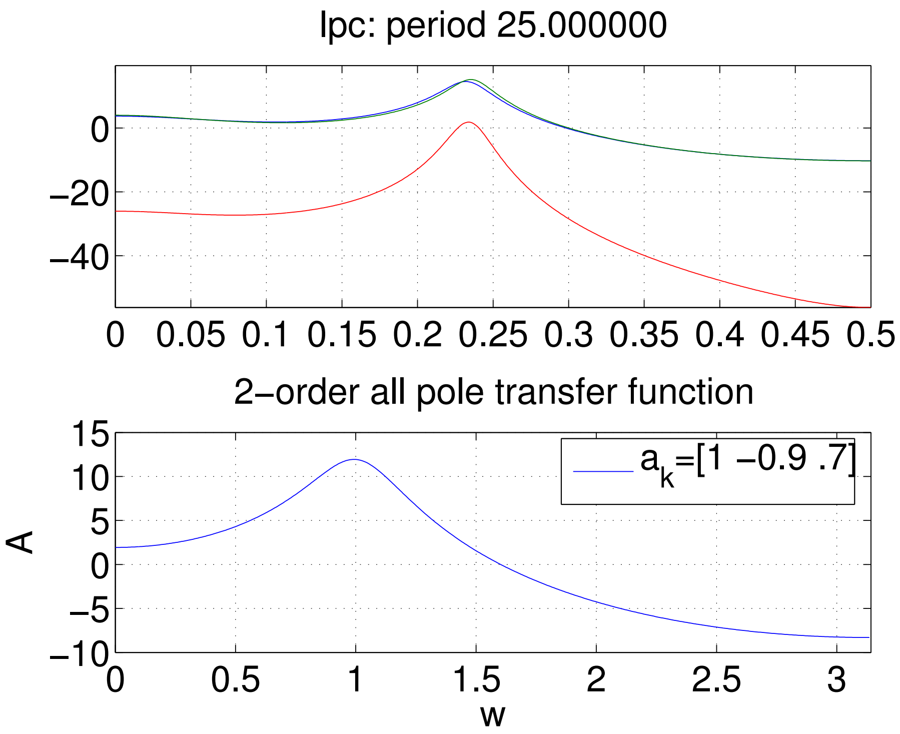


Figure 1: example all pole transfer function of a second order filter

3.1 Estimating the transfer function by means of linear prediction

We suppose the excitation u in eq. (1) is a white noise sequence with uncorrelated samples such that

$$E(u(n)u(n-k)) = \delta(k) \tag{9}$$

where $E(x)$ represents the expected value of x.

- using the impulse response $h(n)$ of the all pole filter to filter the white input signal we get

$$\begin{aligned} s(n) &= (\sum_{k=1}^{P} a_k s(n-k)) + u(n) && (10) \\ s(n) &= u(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k)u(n-k) && (11) \end{aligned}$$

- using eq. (9) we can calculate the energy of the output signal from the linear combination of weighted noise samples. We use $E(u(n-k)u(n-r)) = E(u(n-(k-r)))$

as well as the fact that the expectation operator $E()$ is linear such that we can exchange the order of $E()$ and summation to obtain

$$\begin{aligned}
E(s(n)^2) &= E\left((\sum_{k=-\infty}^{\infty} h(k)u(n-k))^2 \right) && (12)\\
&= \sum_{k=-\infty}^{\infty} E(u(n)^2)h(k)^2 && (13)\\
&\quad +2\sum_{r=-\infty}^{\infty}\sum_{k=r+1}^{\infty} E(u(n)u(n-(k-r)))h(r)h(k) && (14)\\
&= E(u(n)^2)\sum_{k=-\infty}^{\infty} h(k)^2 = \sigma_u^2 \sum_{k=-\infty}^{\infty} h(k)^2 && (15)
\end{aligned}$$

- We conclude that

  $$E(s(n)^2) \geq E(u(n)^2) \tag{16}$$

  with equality only if $h(n) = \delta(n) \rightarrow a_k = 0$ for $k \neq 1$

Knowing that the signal variance increases due to the filter we can search the filter coefficients using the inverse filter

$$\hat{u}(n) = s(n) - \sum_{k=1}^{P} a_k s(n-k) \tag{17}$$

by means of minimizing the variance of the estimated excitation sequence

$$E(\hat{u}(n)^2) = E\left((s(n) - \sum_{k=1}^{P} a_k s(n-k))^2 \right) \tag{18}$$

The solution of this minimization problem leads to a system of linear equations as shown in section **11.1**.

As a result we summarize here that,

- the optimum predictor coefficients will be equal to the system coefficients if the order of the all-pole process and the predictor are the same P and if the excitation signal was white and uncorrelated.
- the optimum predictor of order P is completely defined by the first P samples of the autocorrelation sequence of the process.

3.2 Spectral properties of the LP model

Spectral flatness measure

The flatness of a spectrum $S(w)$ can be measured by means of

$$\xi = \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(w)|^2 dw}}{\frac{1}{2\pi}\int_{-\pi}^{\pi} |S(w)|^2 dw} \tag{19}$$

replacing the integral with approximate summation we find the approximate discrete flatness measure

$$\xi = \frac{e^{\frac{1}{N}\sum_{k=0}^{N-1} \log|S(k)|^2}}{\frac{1}{N}\sum_{k=0}^{N-1} |S(k)|^2} = \frac{\sqrt[N]{\prod_{k=0}^{N-1} |S(k)|^2}}{\frac{1}{N}\sum_{k=0}^{N-1} |S(k)|^2} \tag{20}$$

which shows that the spectral flatness measure is an approximate ratio of geometric and arithmetic mean of discrete spectra.

According to the inequality of the arithmetic and geometric mean for any set of arbitrary postive numbers x we have

$$\sqrt[N]{\prod_{k=0}^{N-1} x_k} \leq \frac{1}{N}\sum_{k=0}^{N-1} x_k \tag{21}$$

with equality only if x_k is constant.

In relation to the spectral flatness measure we conclude that

- ξ is in the range $[0, 1]$.
- $\xi = 1$ only if the spectrum is constant
- with increasing variation in the spectrum ξ decreases
- the minimum value is obtained if part of the spectrum is 0.

If we denote

$$U(w) = S(w)A(w) \tag{22}$$

where S is the source spectrum and A the spectrum of the linear predictor we can investigate the impact of the linear predictor on the flatness of the source spectrum.

- First we investigate the integral over the power density of the residual spectrum $U(w)$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|U(w)|^2)dw \tag{23}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|S(w)|^2|A(w)|^2)dw \tag{24}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|S(w)|^2)dw + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|A(w)|^2)dw \tag{25}$$

and using eq. (8) we find

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|U(w)|^2)dw = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|S(w)|^2dw. \tag{26}$$

- we conclude that the integral of the log power density does not change.

- now we calculate the spectral flatness of the residual

$$\begin{aligned}
\xi_{ee} &= \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|U(w)|^2 dw}}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|U(w)|^2 dw} && (27)\\
&= \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|S(w)|^2 dw}}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|U(w)|^2 dw} && (28)\\
&= \xi_{ss}\frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}|S(w)|^2 dw}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|U(w)|^2 dw} && (29)\\
&= \xi_{ss}\frac{R_S(0)}{R_U(0)} && (30)
\end{aligned}$$

- given the signals energy $R_S(0)$ and its spectral flatness ξ_{ss} it is shown that the residual spectral flatness increases with decreasing residual energy $R_U(0)$.
- result: linear prediction filter creates a **whitening filter**.

Minimum prediction error

Increasing the model order to ∞ will decrease the model error. We now see what is the limiting value

- We may derive the limiting value of the prediction error from the relation between spectral flatness measure of original and residual signal

$$\xi_{ee} = \xi_{ss} \frac{R_S(0)}{R_U(0)}. \tag{31}$$

- and the observation that for $P \rightarrow \infty$ the model spectrum will have exactly matched the signal spectrum, such that the error spectrum is a constant. For this case we know that $\xi_{ee} = 1$.
- we solve for $R_U(0)$

$$\lim_{P\rightarrow\infty} R_U(0) \quad = \quad \xi_{ss} R_S(0) = \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log |S(w)|^2 dw}}{R_s(0)} R_s(0) \tag{32}$$

$$= \quad e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log|S(w)|^2}dw \qquad (33)$$

and find that the minimum prediction error for $P \rightarrow \infty$ is equal to the variation of the input spectrum $S(w)$.

Spectral match

The time domain analysis has shown that for uncorrelated random input signals we may estimate the AR model from the output signal by means of linear prediction.

In the following we investigate into a spectral domain interpretation of LP.

- minimization of the prediction error can be described in the spectral domain. Assume an output signal $x(n)$ and output spectrum $S(w)$ as well as an AR model transfer function $H(w) = \frac{1}{A(w)}$

- The error energy is then

$$\begin{aligned}\sum_n e(n)^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |E(w)|^2 dw = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(w)|^2 |A(w)|^2 dw \qquad (34)\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|S(w)|^2}{|H(w)|^2} dw \qquad (35)\end{aligned}$$

- minimization of the prediction error yields an optimal filter of the given order $H(w)$,
- with that filter we can represent the input spectrum as

$$S(w) = E(w)H(w) = \frac{E(w)}{A(w)} \approx \frac{G}{A(w)}. \qquad (36)$$

  where in the last term we summarize the contribution of the error spectrum as a constant gain factor G.
- the optimal gain factor is determined by the energy of the error sequence

$$G = \sum_n e(n)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(w)|^2 |A(w)|^2 dw \qquad (37)$$

- the representation of the input signal $S(w)$ in terms of the AR model spectrum and a gain factor according to eq. (37) yields a spectral match.

3.3 Model mismatch

In the real cases the correct model order is unknown and the process may not be an AR process.

In these cases it can be shown that:

- increasing the model order will never increase the prediction error σ such that with increasing model order
$$\sigma(P) \geq \sigma(P+1). \tag{38}$$
- for sufficiently large model order the AR model can represent any given transfer function with arbitrary precision.
- the first P coefficients of the autocorrelation function (acf) of the impulse response of the AR model are equal to the first P coefficients of the acf of the signal.

3.4 Error measures

The selection of the *best* AR model among all the different models is controlled by the optimization criterion eq. (18).

There are other optimization criteria that have been proposed for the solution of the inverse filtering problem. All of them can be expressed in terms of the ratio between the signal power density spectrum $|S(w)|^2$ and the model power density spectrum $|\hat{S}(w)|^2$.

$$V(w) = \log(\frac{|S(w)|^2}{|\hat{S}(w)|^2}) \tag{39}$$

Recall LP optimization:

- model power density spectrum estimate is $|\hat{S}(w)|^2 = \frac{G^2}{|A(w)|^2}$
- optimization criterion is minimization of

$$\begin{aligned} E = \sum_n e(n)^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(w)|^2 |A(w)|^2 dw && (40) \\ &= \frac{G^2}{2\pi}\int_{-\pi}^{\pi} \frac{|S(w)|^2}{|\hat{S}(w)|} dw && (41) \\ &= \frac{G^2}{2\pi}\int_{-\pi}^{\pi} e^{V(w)} dw && (42) \\ & && (43) \end{aligned}$$

where ($G = const$).

- never zero, minimum equals excitation energy.
- assumption is white input spectrum

Itakura - Saito measure [MG76] :

- approximate maximum likelihood solution using random Gaussian signal filtered by all pole transfer function. The factor G is not constant but it is optimized as well. U is the residual spectrum. ξ_{ss} is the flatness of the target spectrum S

  Optimization criterion is minimization of

$$I = \int_{-\pi}^{\pi} (e^{V(w)} - V(w) - 1) dw \tag{44}$$

$$= \frac{1}{G^2 2\pi} \int_{-\pi}^{\pi} (|U(w)|^2 - \log(\frac{|S(w)|^2}{|A(w)|^2})) dw - \log(G^2) - 1 \tag{45}$$

$$= \frac{R_U(0)}{G^2} + \log(G^2) - 1 - \log(\xi_{ss} R_S(0)) \tag{46}$$

- AR parameter a_k enter only in $R_U(0)$.
- minimization of I with respect to a_k is equal to minimization of $R_U(0) \longrightarrow$ LP.

- optimization of gain G by means of derivative with respect to G

$$\begin{aligned} 0 &= \frac{\partial}{\partial G}\frac{R_U(0)}{G^2} + \log(G^2) - 1 - \log(\xi_{ss}R_S(0)) && (47)\\ &= -\frac{2R_U(0)}{G^3} + \frac{2}{G} && (48)\\ \rightarrow G &= R_U(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |U(w)|^2 dw && (49) \end{aligned}$$

- assumption is white input spectrum

Mean Squared Error :

- Minimize difference between observed and model spectrum.
- Due to large dynamic range high frequency part of spectrum will be modeled only if log spectrum → cepstrum is used.

$$
\begin{aligned}
O &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(\log(|S(w)|^2) - \log(|\hat{S}(w)|^2)^2 dw && (50)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\frac{|S(w)|^2}{|\hat{S}(w)|^2}\right)^2 dw && (51)
\end{aligned}
$$

- directly related to cepstral smoothing.
- calculates band limited representation of log amplitude spectrum.
- while often used this method is not directly suitable for envelope estimation. It calculates an average energy → mel frequency cepstral coefficients.
- systematic error when used for envelope estimation without further provisions. (see true envelope method below).

3.5 Comparison of Error measures

Impact of $V(w) \neq 0$:

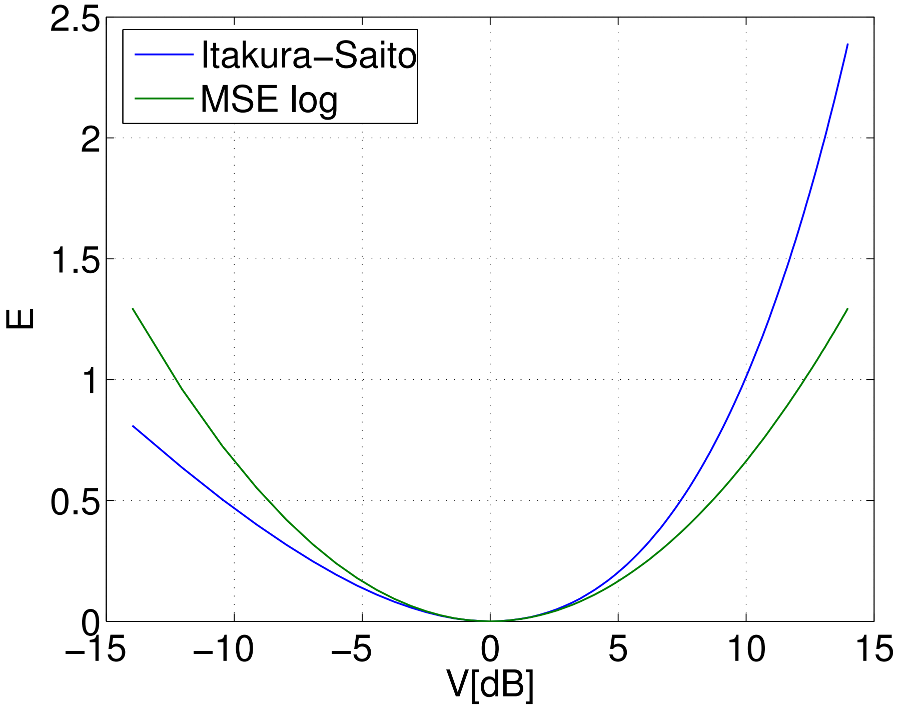


Figure 2: Comparison of error measure for different values of $V(w)$. $V(w) > 0$ source exceeds model, $V(w) < 0$ model exceeds source.

MSE:

- optimization criterion is completely symmetric with respect to sign of error.
- best for envelope manipulation because under and over estimation of filter transfer function will on average be the same.
- if envelope shall be moved the errors in the low and high energy parts of the spectrum have the same effect, so the MSE is appropriate.

Itakura-Saito measure/LP measure:

- these two measures yield essentially the same solutions, therefore, they apply the same asymmetric evaluation.
- overestimation of the spectrum is much less significant then underestimation.
- If the model spectrum cannot fit the signal spectrum the model will tend to follow the peaks and shortcut the valleys.
- better suited for formant estimation because high amplitude regions (formants) are better represented then valleys.

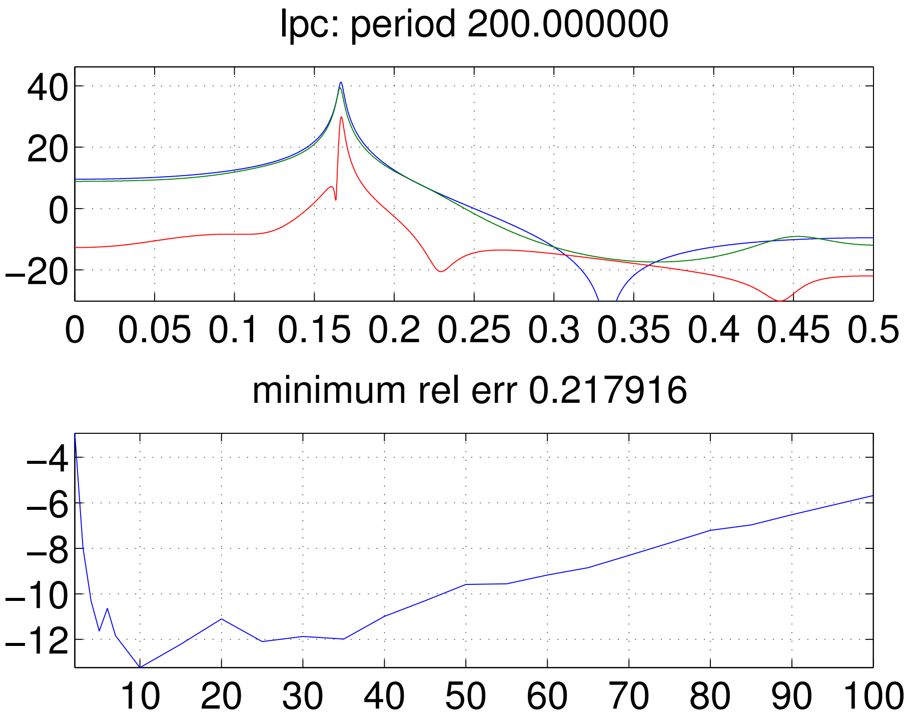


Figure 3: log amplitude error between original filter transfer function with $P = 2$ poles and $Z = 2$ zeroes and the best estimated filter transfer function using lp predictor and different orders (top) and log amplitude error as a function of the model order (bottom).

4 Envelope estimation from harmonic spectra

If the input sequence u is not white the linear prediction will try to whiten the spectrum of the input signal as well

- for sufficiently low fundamental frequency (compared to the format band with of the target filter) the linear predictor will work correctly.
- especially problematic for high pitched harmonic sounds
- linear predictor will try to whiten the harmonics
- systematic error when estimating of the system transfer function is desired.

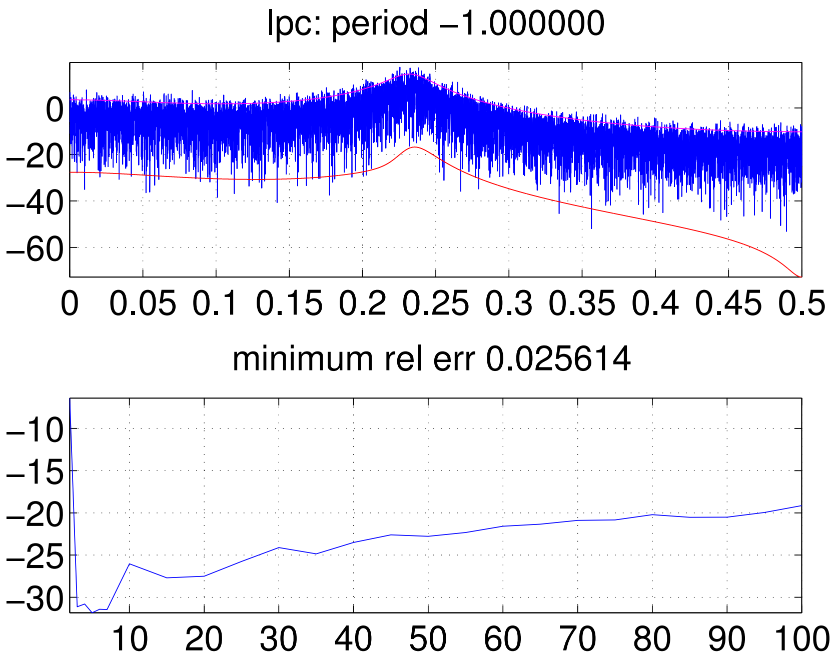


Figure 4: Linear prediction and spectral estimation. Problem is the estimation of an unknown all pole filter with order P=3 and white noise excitation signal. The input spectrum target (blue), transfer function (green) and estimation error for the best estimate (red) are displayed on top. The bottom graph shows the relation between model order and the error of the estimated transfer function. Optimal model filter performance is achieved for $P_m \in 3, 4, 5$

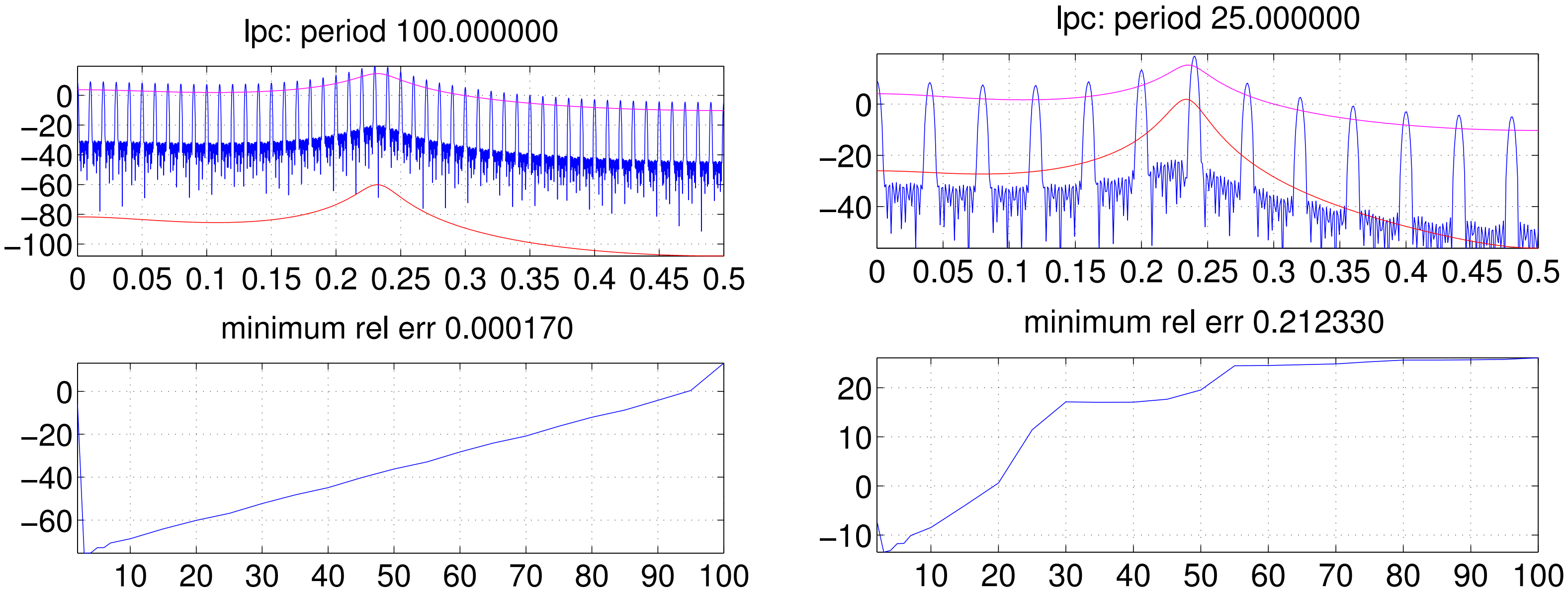


Figure 5: Linear prediction and spectral estimation. Problem is the estimation of an unknown all pole filter with order P=3 and a periodic input signal. The input spectrum target (blue), transfer function (green) and estimation error for the best estimate (red) are displayed on top. The bottom graph shows the relation between model order and the error of the estimated transfer function. The (normalized) fundamental frequency is 1/100 (left) and 1/25 (right). The optimal model order is $P_m = 3$ but the minimum spectral error is much smaller for the lower fundamental frequency.

5 Discrete all pole (DAP)

How to reduce the systematic error of linear prediction?

- the optimal linear predictor is completely defined through the first P samples of the auto correlation sequence of the process.
- for noise signals the akf of the filter and the expectation value of the acf of the filtered signal are the same.

  Suppose an uncorrelated white noise sequence $x(n)$ with acf $\delta(n)$ and an AR model with transfer function $A(w)$ and impulse response $h(n)$

  The signal $s(n) = x(n) * h(n)$ has a spectrum

$$S(W) = X(w)A(w) \tag{52}$$

By means of the inverse Fourier transform of the signal spectrum we can calculate the acf of the filtered signal. Because the acf is the IDFT of the squared magnitude spectrum the acf of the signal $R_s(n)$ is equal to

$$R_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}(|X(w)||A(w)|)^2 dw \tag{53}$$

and because the expectation of the noise signal spectrum is a constant the average or expected acf becomes

$$E(R_s(n)) = E(\frac{1}{2\pi}\int_{-\pi}^{\pi}(|X(w)||A(w)|)^2 dw) = \sigma^2 E(\frac{1}{2\pi}\int_{-\pi}^{\pi}(|A(w)|)^2 dw) \tag{54}$$

which, besides the scaling factor σ^2 is equal to the acf of the filter.

- therefore, an all pole filter can be estimated without systematic error from a filtered white noise sequence.

For an harmonic signal the relations are less advantageous.

- for the power spectrum of the transfer function $|S(w)|^2$ and the related auto correlation sequence $R_{ori}(n)$ we have the following relation

$$|S(w)|^2 = \sum_{n=-\infty}^{\infty} R_{ori}(n) e^{-jwn} \tag{55}$$

- ideal sampling of the power spectrum by means of harmonic partials with fundamental frequency w_0 creates a new signal $s(n)$ with auto correlation sequence $R_s(n)$ equal to

$$R_s(n) = \frac{1}{N} \sum_{m=-M}^{M} |S(w_0 m)|^2 e^{jw_0 mn} \tag{56}$$

- by means of inserting eq. (55) into eq. (56) we get the relation between the original and the sub sampled acf

$$R_s(n) \quad = \quad \sum_{m=-M}^{M} \sum_{l=-\infty}^{\infty} R_{ori}(l) e^{jw_0 m(n-l)} \tag{57}$$

$$= \sum_{l=-\infty}^{\infty} R_{ori}(l) \frac{1}{N} \sum_{m=-M}^{M} e^{jw_0 m(n-l)} \tag{58}$$

- we investigate the second term of the convolution

$$\sum_{m=-M}^{M} e^{jw_0 mn} = \sum_{m=0}^{2M} e^{jw_0 (m-M)n} \tag{59}$$

$$= e^{-jw_0 Mn} \sum_{m=0}^{2M} e^{jw_0 mn} \tag{60}$$

$$= e^{-jw_0 Mn} \frac{1 - e^{(2M+1)w_0 n}}{1 - e^{w_0 n}} \tag{61}$$

$$= \frac{e^{-jw_0 (M+\frac{1}{2})n}}{e^{-jw_0 \frac{1}{2} n}} \frac{1 - e^{(2M+1)w_0 n}}{1 - e^{(w_0 n}} \tag{62}$$

$$= \frac{\sin\,(M+\frac{1}{2})w_0 n}{\sin\frac{1}{2}w_0 n} \tag{63}$$

- periodic sampling by means of the ideal harmonic peak spectrum yields a convolution of the auto correlation function with a periodic sinc function of period $\frac{2\pi}{w_0}$.

$$R_s(n) = \sum_{l=-\infty}^{\infty} R_{ori}(l)\frac{1}{N}\frac{\sin((M+\frac{1}{2})w_0(n-l))}{\sin(\frac{1}{2}w_0(n-l))} \tag{64}$$

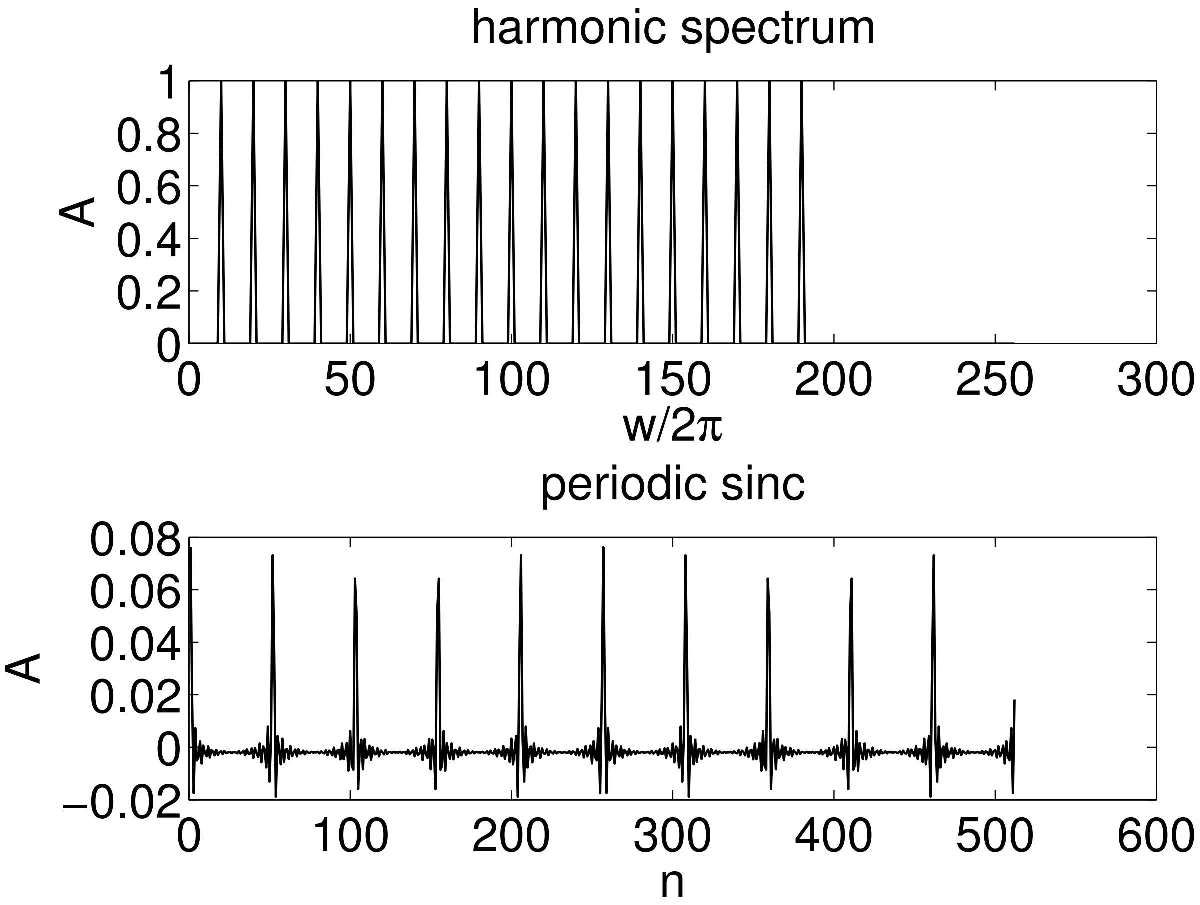


Figure 6: periodic sinc function for a fundamental frequency of $w_0 = \frac{2\pi}{51}$

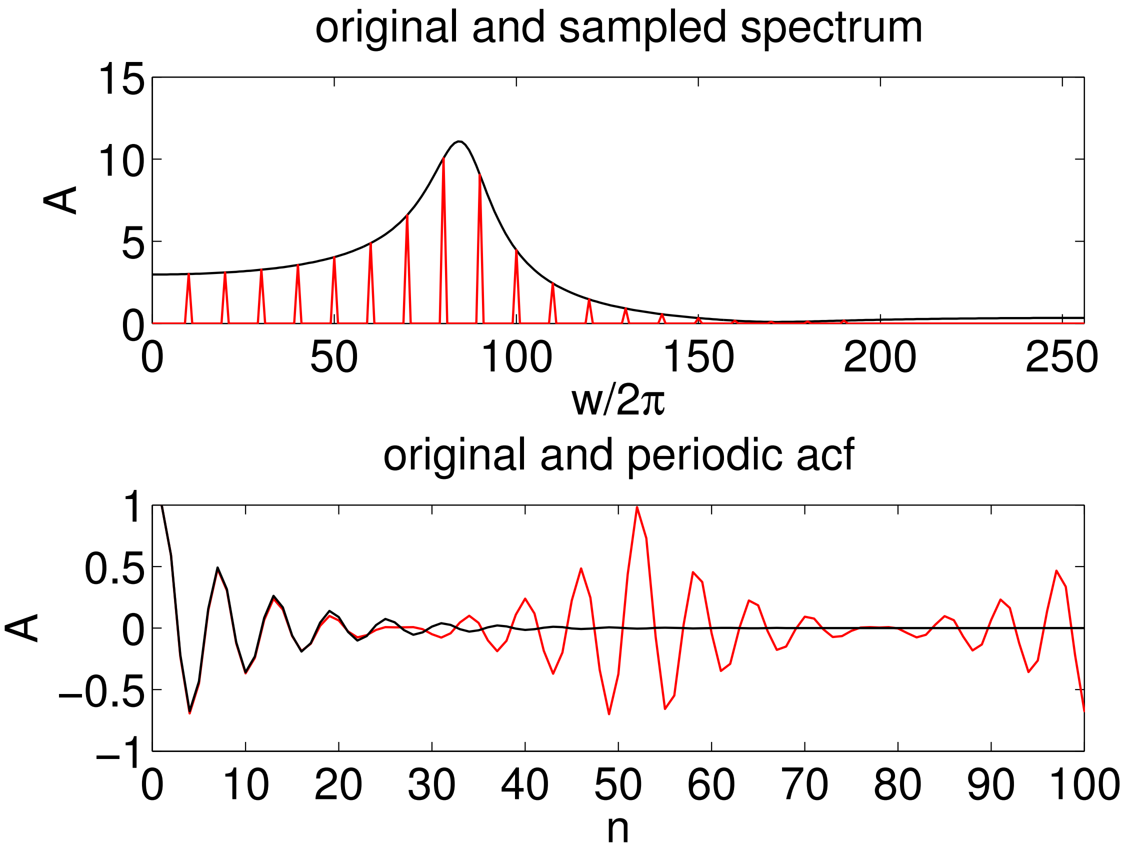


Figure 7: comparing original and periodicially convolved acf

- the subsampling in the spectral domain creates a periodic repetition in the time domain.
- if $R_{ori}(l)$ is not time limited the periodic repetition creates time aliasing which one source of the problems when estimating the AR model.

The discrete all pole algorithm uses a discrete version of the Itakura-Saito measure to adapt the aliased acf of the all-pole model to the aliased acf of the signal

Objective function

$$E_{IS} = \sum_{m=0}^{M} \frac{|S(w_m)|^2}{|\hat{S}(w_m)|^2} - \log(\frac{|S(w_m)|^2}{|\hat{S}(w_)|^2}) - 1 \quad (65)$$

The systematic errors disappear because of the fact that system and model transfer function are subject to the same sampling such that both acf that will be used will contain the same aliasing.

advantages :

- removal of a large part of the systematic errors,

- works correctly for noise spectra and harmonic spectra, not as good for spectra with frequency dependent SNR.
- equivalent to LP model if number of peaks is sufficiently large.

disadvantages :

- optimal order difficult to determine,
- iterative procedure is rather slow,
- requires peak picking.

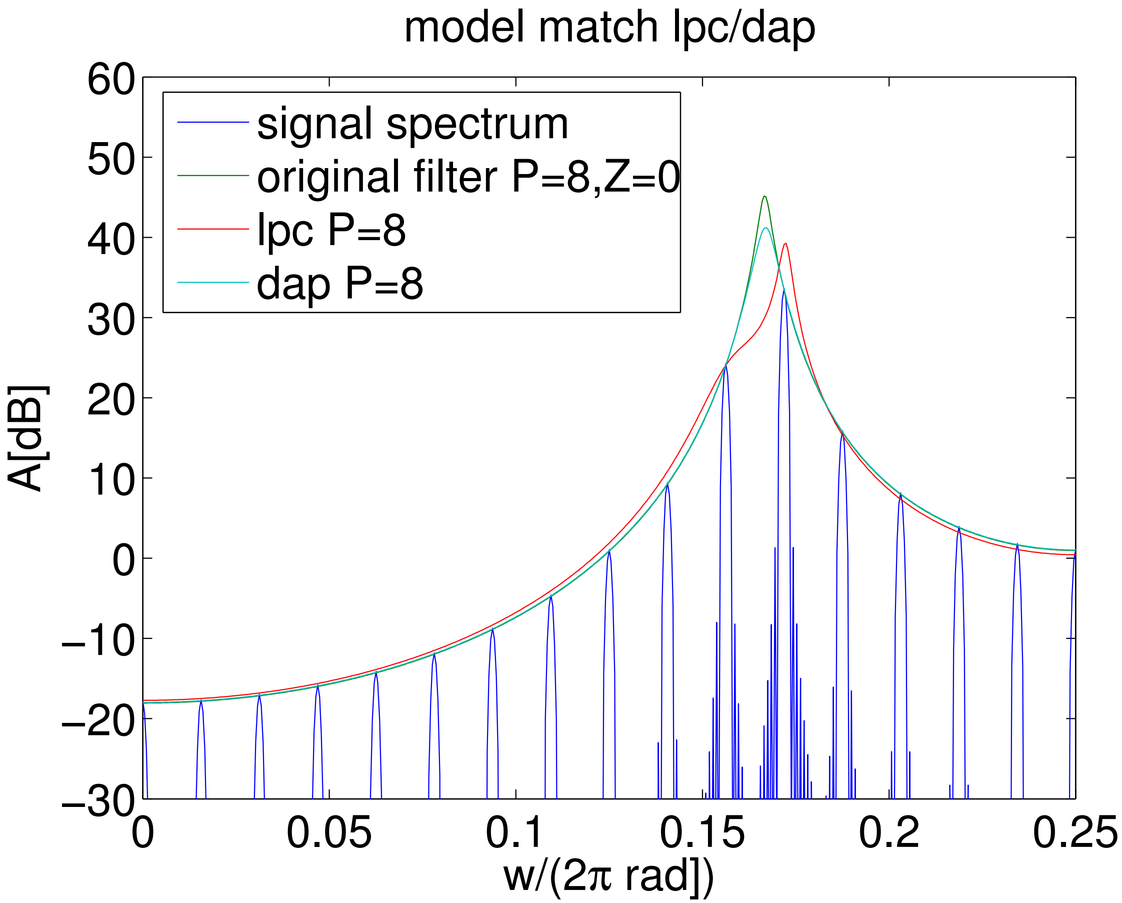


Figure 8: comparing lp and dap filter estimated from a periodic sequence with all pole envelope of order $P = 8$ using model orders $P_m = P$.

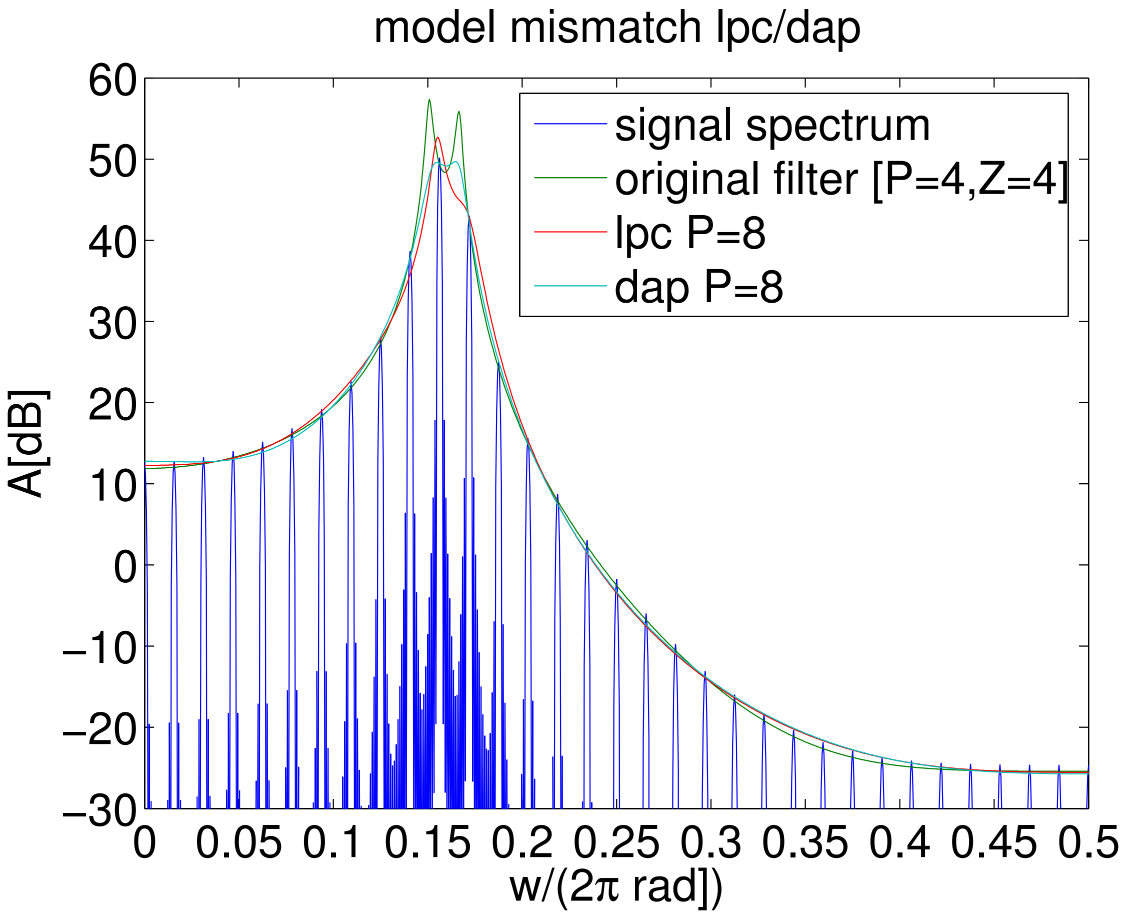


Figure 9: comparing lp and dap filter estimated from a periodic sequence

6 The Cepstrum

There is an alternative approach to envelope estimation of harmonic signals based on cepstral smoothing.

- Given the signal $x(n)$ with Fourier spectrum $X(w)$ we define the real cepstrum as the inverse DFT of the log amplitude spectrum

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(|X(w)|) e^{jwn} dw \tag{66}$$

- note that due to the even symmetry $|X(w)| = |X(-w)|$ for for real sequences $x(n)$ the cepstral coefficients are always real.
- the cepstrum is an important tool in digital signal processing. One of the initial reasons to define the cepstrum was the fact that due to the log function the spectral multiplication would be transformed into sum. Assume a filter $H(k)$ that is applied to a source

signal s with spectrum $S(k)$. For the cepstrum we get

$$
\begin{aligned}
c(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|H(k)||S(k)|)e^{jwn}dw && (67)\\
c(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(|H(k)|)+\log(|S(k)|)e^{jwn}dw && (68)\\
c(n) &= \frac{1}{2\pi}(\int_{-\pi}^{\pi}\log(|H(k)|)e^{jwn}dw)+\int_{-\pi}^{\pi}\log(|S(k)|)e^{jwn}dw && (69)\\
c(n) &= c_s(n)+c_h(n) && (70)
\end{aligned}
$$

the cepstrum of the filtered signal is just the sum of the cepstra of the original signal and the filter.

- Due to symmetry of the log amplitude spectrum the cepstral coefficients $c(n)$ are always real.

  Another important consequence of the log amplitude spectrum is the fact that the poles and zeros of rational filters each contribute by means of adding a decaying exponential to the cepstrum of the transfer function. For a proof see [Smi05].

- A minimum phase pole or zero of a filter (a pole or zero inside the unit circle) the related decaying exponential is casual. This means $c(n) = 0$ for $n < 0$.
- A maximum phase pole or zero of a filter (a pole or zero outside the unit circle) the related decaying exponential is anti-causal. This means $c(n) = 0$ for $n > 0$.
- creating minimum phase from zero phase by simply swapping the anti-causal part into the causal part.

6.1 cepstral smoothing

in the context of envelope estimation we are interested to find a smooth curve that connects the spectral peaks.

cepstral smoothing :

- uses only the $L+1$ low order coefficients of the cepstrum,
- similar to low pass filtering a signal by means of taking only the low order coefficients of the spectrum,
- as is shown in [Röb06, section: 2.3], the Fourier coefficients minimize the residual squared error. Because the target spectrum is the log amplitude spectrum, cepstral smoothing will create an approximation of the log amplitude spectrum that minimizes

$$E = \sum_k |S(k) - \sum_{n=0}^{L} c_n e^{-j\frac{2\pi}{N}kn}|^2 \qquad (71)$$

- L is called the order of the cepstral smoothing filter

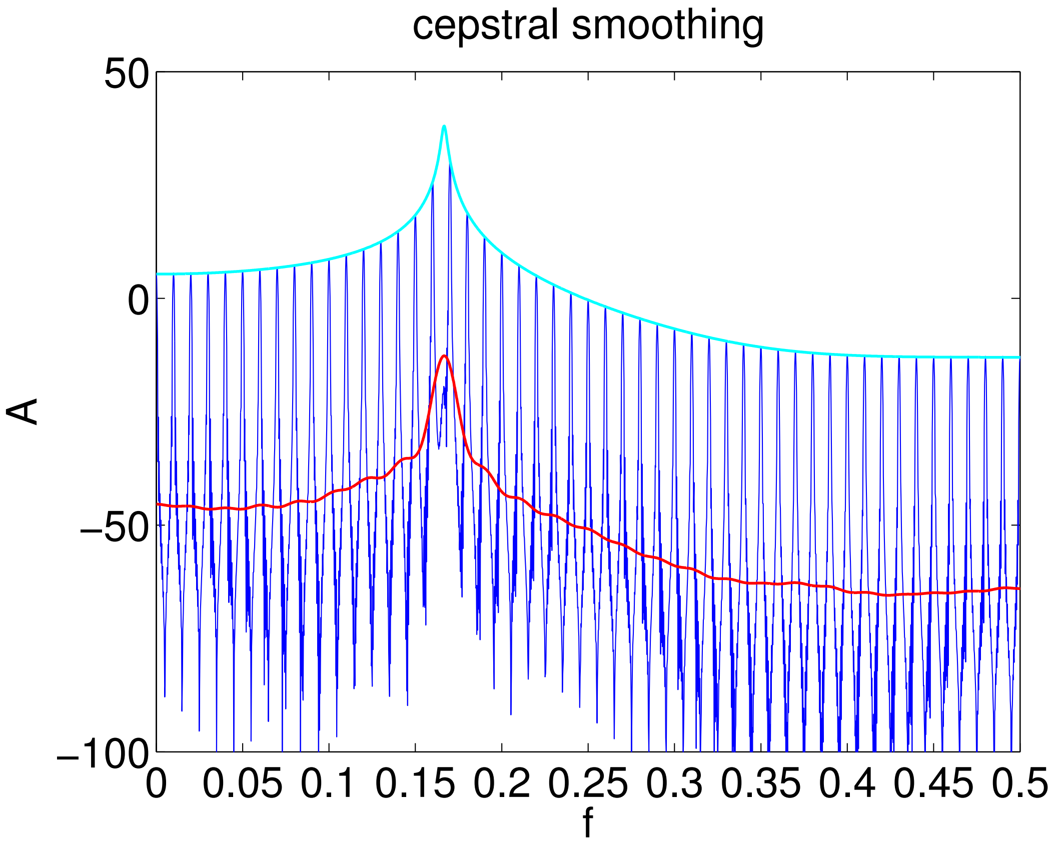


Figure 10: The result of the cepstral smoothing of the log amplitude spectrum with a given envelope for smoothing order $L = 50$ does not match well with the filter shown as spectral envelope of the spectrum.

7 Envelope estimation using the cepstrum

Discrete Cepstrum: [CLM95, Oud97]

- select interesting peaks
- solve minimum error problem for the amplitude frequency pairs related to the selected peaks,
- due to irregular sampling the sinusoidal basis functions are no longer orthogonal and the a costly matrix inversion has to be performed. different bas

True envelope: [IA79, RR05]

- **Iteratively updated cepstrum**
- Originally proposed in **Japan** in 1970s,
- iteratively apply cepstral smoothing and then produce the maximum of the original and the spectrally smoothed spectrum until the whole spectrum is covered.

Examples:

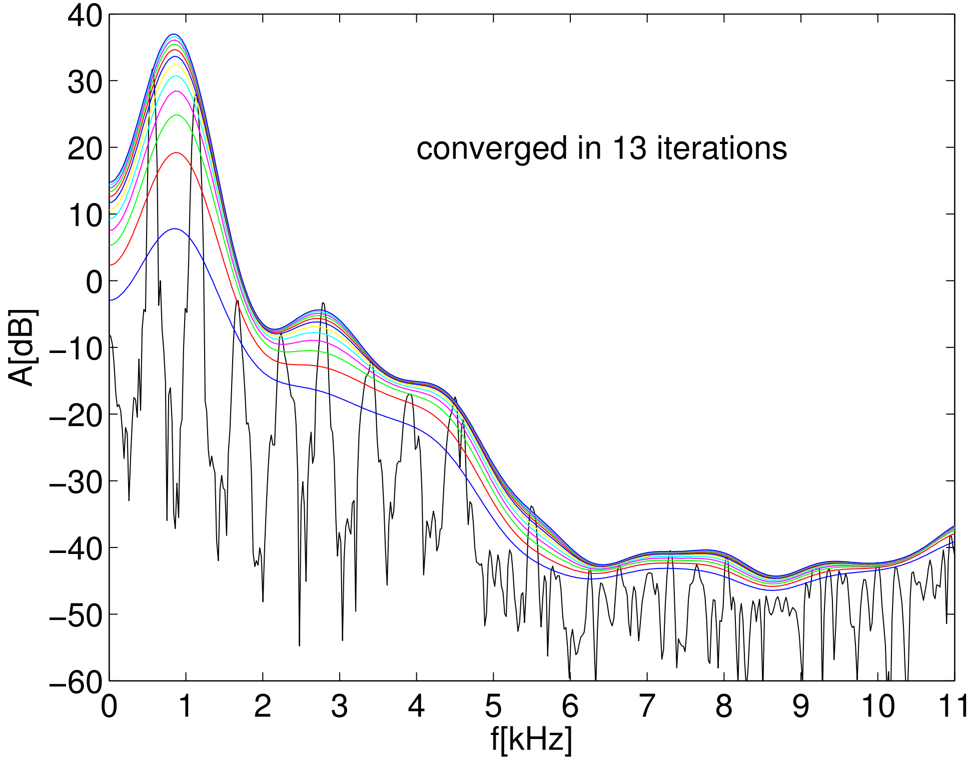


Figure 11: true envelope estimator applied to a high pitched singing voice amplitude spectrum.

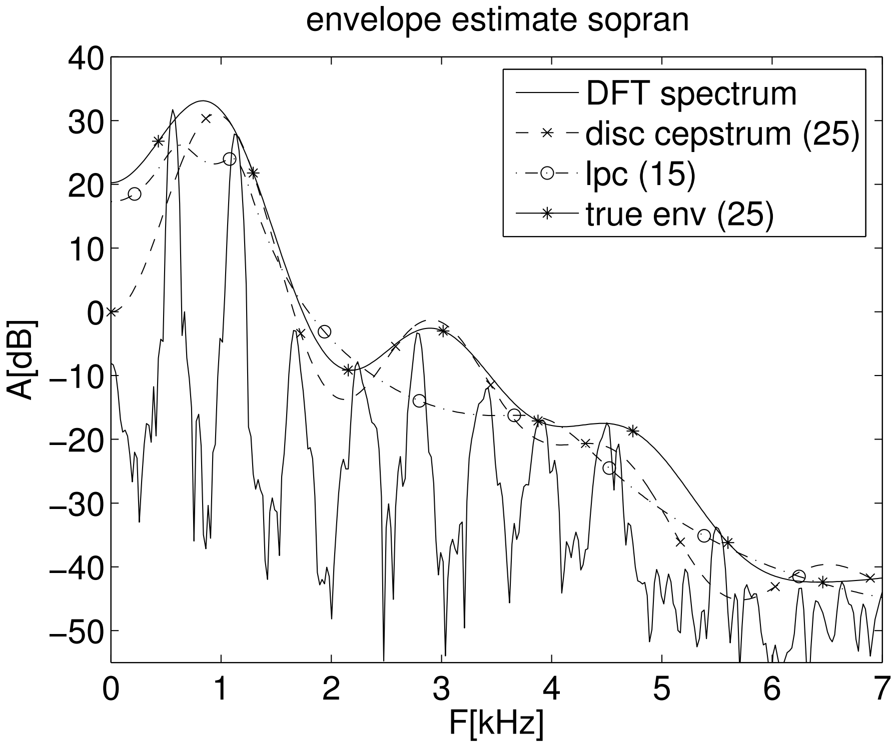


Figure 12: comparison of LP, discrete cepstrum and true envelope estimates applied to a high pitched singing voice envelope estimation. Discrete cepstrum reacts less robust then true envelope with respect to the large hole around 0Hz, moreover did the peak selection algorithm miss some peaks. LPC misses the singer formant around 3kHz and creates to many details for the two fundamental peaks of the spectrum.

7.1 True Envelope and Order selection

The order selection problem has been discussed for the all pole models.

Order selection is less problematic for the cepstral interpolation methods.

- **Observed spectrum** contains information about the **resonator filter** or the **spectral envelop** only at the **prominent spectral peaks**.
- Due to **spectral sub sampling** of the transfer function only **band limited envelopes** can be estimated.
- time domain sampleperiode $T = \frac{1}{S_R}$ translates into Nyquist bandlimit $0.5S_R$.
- Imagine the spectral envelope is the original function which is sampled with a sampleperiod of $\frac{2\pi}{N}$ and the related time bandwidth is N and the highest order of coefficients $0.5N$.
- If the envelope is sampled by means of an harmonic comb of “period” $m\frac{2\pi}{N}$ then the sampleperiod is multiplied by m because the original sample period in frequency domain was $\frac{2\pi}{N}$. We conclude that the timewidth in cepstral domain is $\frac{N}{m}$ and the highest bin that can be used to represent the band limited data without suffering from aliasing

due to subsampling in the frequency domain is $\frac{N}{m}$.

$$O_T(F_0) = \frac{0.5SR}{F_0}$$

- For some instruments the distance δ_f between the peaks carrying **envelope information** is larger than $f0$ (Clarinet $\delta_f = 2f0$) and the limiting order is then $O_T(\delta_f)$.

8 All-pole and cepstral parameter conversion

There exists a direct transformation from the all pole coefficients a_k into the cepstral coefficients c_k [Ata74].

We start with the definition of the cepstrum and using the fact that $A(w)$ is minimum phase such that the cepstrum will be causal

$$\log(\frac{G}{A(w)}) = \sum_{k=0}^{\infty} c_k e^{-jwk} \tag{72}$$

$$\log(\frac{G}{1-\sum_{k=1}^{M} a_k e^{-jwk}}) = \sum_{k=0}^{\infty} c_k e^{-jwk} \tag{73}$$

$$\log(G) - \log(1-\sum_{k=1}^{M} a_k e^{-jwk}) = \sum_{k=0}^{\infty} c_k e^{-jwk} \tag{74}$$

differentiation of both sides with respect to e^{-jw} and then multiplication by the same facor yields

$$\frac{\sum_{k=1}^{M} k a_k e^{-jw(k-1)}}{1 - \sum_{k=1}^{M} a_k e^{-jwk}} = \sum_{k=1}^{\infty} k c_k e^{-jw(k-1)} \tag{75}$$

$$\sum_{k=1}^{M} k a_k e^{-jwk} = (1 - \sum_{k=1}^{M} a_k e^{-jwk}) \sum_{k=1}^{\infty} k c_k e^{-jwk} \tag{76}$$

$$\tag{77}$$

Now substitute $e^{jw} = z$ and equate the factors belonging to the same polynomial order z^{-k}

$$\sum_{k=1}^{M} k a_k z^{-k} = (1 - \sum_{k=1}^{M} a_k z^{-k}) \sum_{k=1}^{\infty} k c_k z^{-k} \tag{78}$$

the equation of the factors of z^{-1} is

$$c_1 = a_1 \tag{79}$$

for $k = 2$ we get

$$c_2 = 0.5 a_1 c_1 + a_2 \tag{80}$$

and in continuing one may find a recursive relationship between the a_k and c_k

$$c_n = \sum_{k=1}^{n-1} (1 - \frac{k}{n}) a_k c_{n-k} + a_n \quad \text{with} \quad a_k = 0 \quad \forall \quad n > P \tag{81}$$

which allows to calculate a cepstral representation of the filter transfer function directly from the predictor coefficients. The equation can be solved for a_n as well which results in

$$a_n = -\sum_{k=1}^{n-1} (1 - \frac{k}{n}) a_k c_{n-k} + c_n \tag{82}$$

such that the predictor model of arbitrary order can be derived from the cepstral coefficients.

c_0 can be obtained from the definition using eq. (8)

$$
\begin{aligned}
c_0 &= \frac{1}{2\pi}\int_{\pi}^{\pi}\log(\frac{G}{|A(w)|})dw & (83)\\
&= \frac{1}{2\pi}\int_{\pi}^{\pi}\log(G)-\log(|A(w)|)dw = \log(G) & (84)
\end{aligned}
$$

Note, that the conversions above are obtained under the assumption that the cepstral and all-pole model have infinite order. The conversion from LPC to cepstral coefficients using the above formulas can be understood as a polynomial or Taylor approximation of the log amplitude transfer function of the all pole system. While this conversion is often used to obtain a finite number of cepstral coefficients from all pole models to classify speech signals, the MSE of the finite approximation obtained with the above conversion is certainly not the minimum error that can be obtained for the given cepstral order.

9 Comparison of the envelope models

Having seen so many different approaches to achieve envelope estimation it is interesting to ask which model is the best.

As always there is no unique answer.

Experiments:

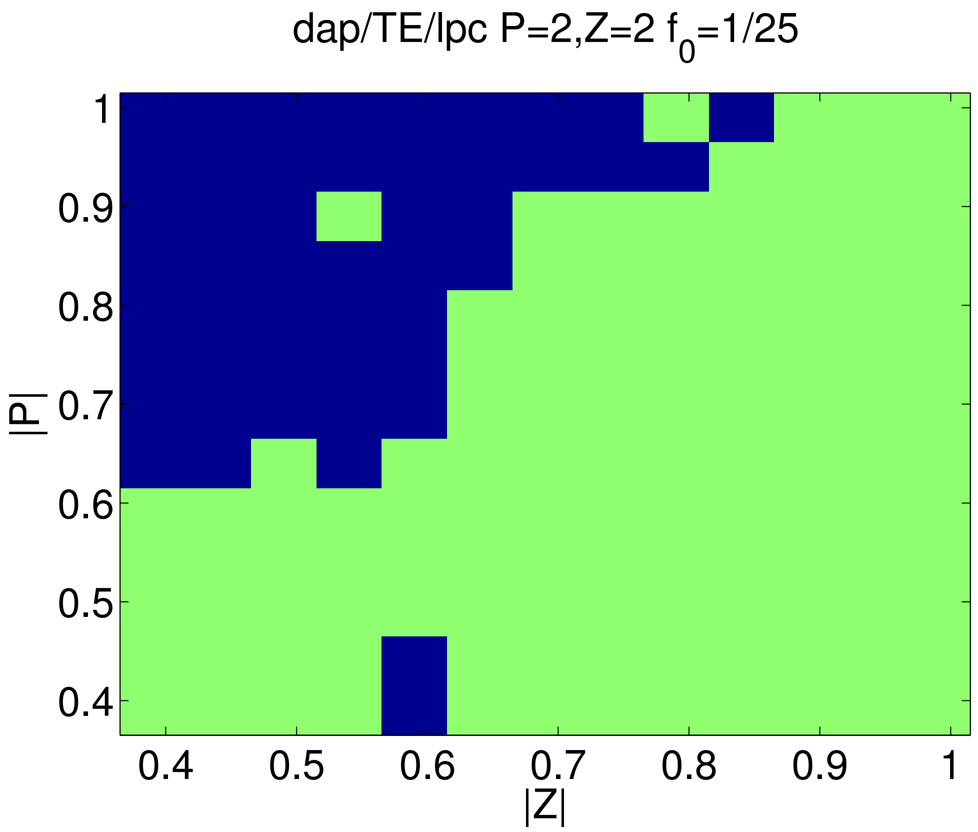


Figure 13: Comparing LP, DAP and TE estimator using minimum squared error of log in relation to pole and zero position. Optimum model is color coded: blue =DAP, red=LP and green = TE.

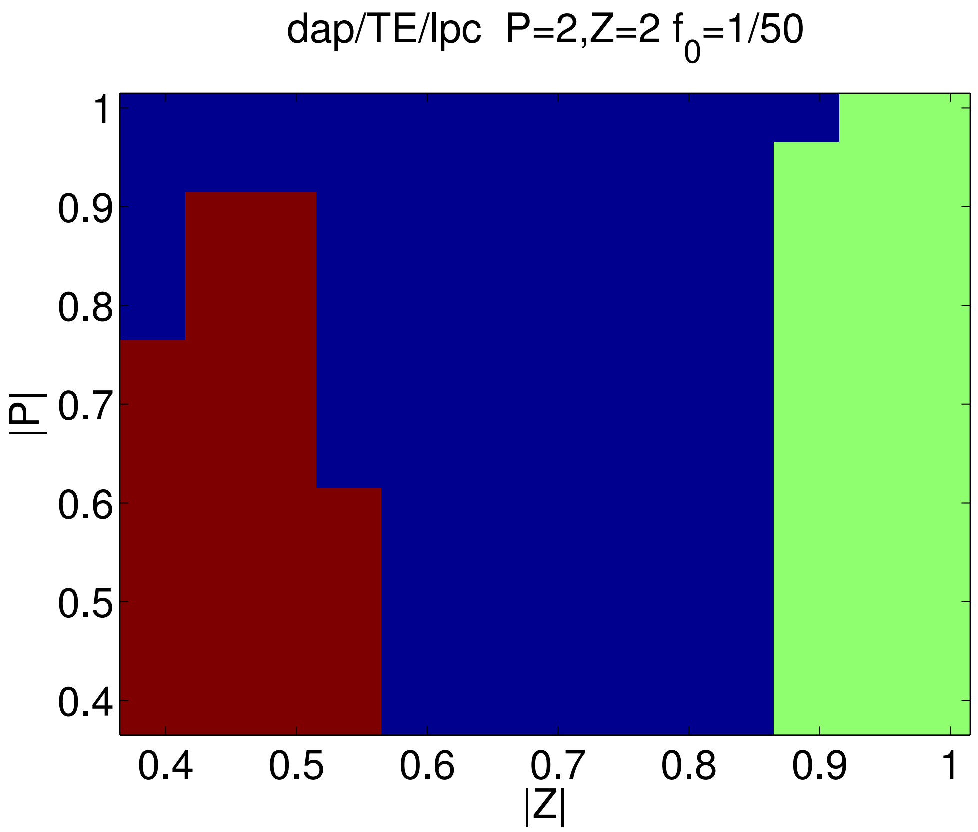


Figure 14: comparing LP, DAP and TE estimator for minimum squared error in relation to pole and zero position. optimum model is color coded: blue =DAP, red=LP and green = TE.

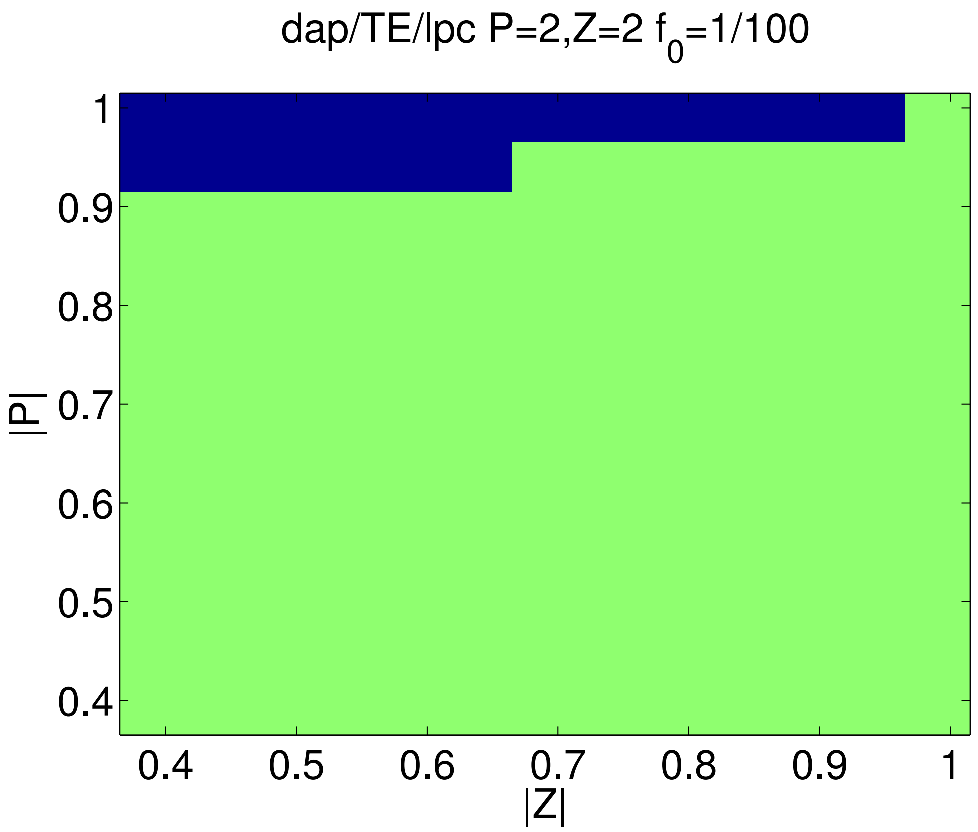


Figure 15: comparing LP, DAP and TE estimator for minimum squared error in relation to pole and zero position. optimum model is color coded: blue =DAP, red=LP and green = TE.

10 Application of envelope models for signal modification

Due to the direct link between sound timbre and envelope the estimated spectral envelopes can be used for a large number of signal transformations:

- separately transpose pitch and timbre of a sound,
- remove timbre information by filtering with inverse envelope to obtain the white excitation signal,
- use spectral envelope of one sound to filter a sound or its excitation signal,
- scale the spectral envelope to enhance/attenuate the timbre.

11 Appendix

11.1 Linear prediction

We want to minimize eq. (18) the excitation variance of the predictor

$$\hat{s}(n) = \sum_{k=1}^{P} a_k s(n-k) \tag{85}$$

so we search the solution to the problem

$$\min_{a_k} = E(e(n)^2) = E\left((s(n) - \hat{s}(n))^2\right) \tag{86}$$

setting the differentiation with respect to each parameter a_k to zero yields

$$\begin{aligned} 0 &= \frac{\partial}{\partial a_k} E\left((s(n) - \hat{s}(n))^2\right) && (87) \\ &= -2E\left(s(n) - \hat{s}(n))s(n-k)\right)) && (88) \end{aligned}$$

$$= \quad R(k) - \sum_{l=1}^{P} a_l R(k-l) \quad \text{for} \quad k = 1, \ldots, P \tag{89}$$

where $R(k)$ is the auto correlation sequence

$$R(k) = E(s(n)s(n-k)) \tag{90}$$

These equations can be combined into a system of linear equations

$$\begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(P) \end{pmatrix} = \begin{pmatrix} R(0) & R(-1) & \ldots & R(-P+1) \\ R(1) & R(0) & \ldots & R(-P+2) \\ \vdots & & \ddots & \vdots \\ R(P-1) & R(P-2) & \ldots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{pmatrix} . \tag{91}$$

The solution gives the optimal predictor coefficients $\boldsymbol{a} = a_1, \ldots, a_P$

We can make the following conclusions.

- If the generating process was an all pole system of the same order P with coefficients $\boldsymbol{\alpha}$ and the excitation sequence was uncorrelated then $\boldsymbol{\alpha} = \boldsymbol{a}$ and the error $e(n)$ will be equal to the excitation sequence $u(n)$.

  This is due to the fact that any mismatch between the filter coefficients would increase the error variance.

- the minimum error is

$$\begin{aligned}
\sigma &= E\left((s(n) - \hat{s}(n))^2\right) && (92)\\
&= E\left((s(n) - \hat{s}(n))(s(n) - \sum_{k=1}^{P} a_k s(n-k))\right) && (93)\\
&= E\left((s(n) - \hat{s}(n))s(n)\right) && (94)\\
&= E\left(s(n)s(n) - \sum_{k=1}^{P} a_k s(n-k)s(n)\right) && (95)
\end{aligned}$$

$$= \quad R(0) - \sum_{k=1}^{P} a_k R(k) \tag{96}$$

where we have used eq. (88).

References

[Ata74] B. S. Atal. Effectiveness of linear prediction characteristiucs of the spech wave for automatic speaker identification and verification. *Jour. Acoutic Soc AM*, 55(6):1304–1312, 1974. 59

[CLM95] O. Cappé, J. Laroche, and E. Moulines. Regularized estimation of cepstrum envelope from discrete frequency points. In *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1995. 54

[IA79] S. Imai and Y. Abe. Spectral envelope extraction by improved cepstral method. *Electron. and Commun. in Japan*, 62-A(4):10–17, 1979. in Japanese. 54

[MG76] J. D. Markel and A. H. Gray. *Linear Prediction of Speech*. Springer Verlag, 1976. 11, 28

[Oud97] Marine Oudot. Estimation of the spectral envelope of mixed spectrum signals using a penalized likelihood criterion. *IEEE Transactions on Speech and Audio Processing*, 1997. 54

[Röb06] A. Röbel. Analysis, modelling and transformation of audio signals - Part I: Fundamentals of discrete fourier analysis. lecture slides, 2006. AMT : Part I. 52

[RR05] A. Röbel and X. Rodet. Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation. In *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx05)*, pages 30–35, 2005. 54

[Smi05] Julius O. Smith. *Introduction to Digital Filters, September 2005 Draft*. `http://ccrma.stanford.edu/˜jos/filters05/`, May 2005. see: section `Poles_Zeros_Cepstrum.html`. 50