Minimum Phase Compensation in Speech Coding using Hammerstein Model

Jari Turunen, Juha T. Tanttu, Frank Cameron
Tampere University of Technology, Pori
Pohjoisranta 11, P.O. Box 300, FIN-28101, Pori, Finland
{jari.j.Turunen, juha.tanttu, frank.cameron}@tut.fi

Abstract
When using Bai’s method for estimating the coefficients of both the linear and nonlinear parts of a Hammerstein model, there is no guarantee that the resulting linear filter will be stable. To obtain a stable linear filter the well-known minimum phase correction can be used. However, if it is used without some compensation in the nonlinear part, one obtains a Hammerstein model with no theoretical foundation. In this paper we present a simple method for compensating the nonlinear part for corrections performed to the linear part. The compensation proved to be stable and fulfills the requirements of combined parameter estimation.

1. Introduction
Nonlinear speech processing has been an active research area for several decades. Several different approaches have been presented during the years: fluid dynamics, chaotic modeling, neural net based estimators [1-7] etc. A Hammerstein model is one way to model speech using a static nonlinearity combined with a linear filter. The availability of an efficient coefficient identification process [8, 9] makes using the Hammerstein model attractive for time series modeling, such as speech modeling for coding purposes [10, 11]. This identification process estimates the coefficients of both the linear and nonlinear parts together using a fast, noniterative method, however it does not ensure stability for the linear filter. The LPC modeling approach, that is the commonly used method in speech processing, guarantees stable linear filter coefficients. We propose using Bai’s identification method in conjunction with the minimum phase correction, which is a well-known way to transform an unstable linear filter to a stable one, without altering the gain properties of the filter. However, if the minimum phase correction is simply applied to the linear filter while leaving the coefficients of the nonlinear part unaltered, the resulting Hammerstein model has no theoretical foundation. In this paper we present a simple way to modify the nonlinear coefficients after the minimum phase correction has been applied.

2. Hammerstein Model
The Hammerstein model, in Fig. 1, is a combined static nonlinearity followed by a linear time-invariant system. This structure can be imagined also as one-step predictor as in [12]:

\[ \hat{u}(n) = \sum_{k=1}^{p} b_k \sum_{r=0}^{r+1} a_r g_r(u(n-k)) + w(n) \]  

(1)

where the \( a_i, i = 0 \ldots r \), are the unknown nonlinear coefficients, \( g_i \) is a nonlinear function and \( r+1 \) is the number of nonlinear functions and coefficients. In this paper the linearity will be assumed to be an AR model; for the case of a general linear model see [1]. The linear AR-filter coefficients are denoted by \( b_k, k=1 \ldots p \). The modeling error is \( w(n) \).

We next describe how the coefficients of the nonlinear and linear parts can be estimated. Equation (1) can be written as:

\[ \hat{u}(n) = \sum_{k=1}^{r} b_k a_r g_r(u(n-k)) + w(n) \]  

(2)

This in turn can be written as

\[ \hat{u}(n) = \theta^T \phi + w(n) \]  

(3)

where the parameter vector \( \theta \) and the data vector \( \phi \) are

\[ \theta = [h_0 a_0, \ldots, h_0 a_r, \ldots, h_r a_0, \ldots, h_r a_r, \hat{y}]^T \]  

(4)

\[ \phi = [g_0(u(n)), \ldots, g_r(u(n)), \ldots, g_0(u(n-p)), \ldots, g_r(u(n-p))]^T \]  

(5)

When N samples have been collected, the following can be formed:

\[ \hat{U}_N = \Phi_N \theta + W_N \]  

(6)

\[ \hat{U}_N = [y_1, y_2, \ldots, y_N]^T \]  

\[ \Phi_N = [\phi_1, \phi_2, \ldots, \phi_N]^T \]  

(7)

\[ W_N = [w_1, w_2, \ldots, w_N]^T \]  

The least squares estimate of \( \theta \) is given by:

\[ \hat{\theta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N U_N \]  

(8)
The parameter vector $\theta$ from equation (8), can be organized into block column matrix:
\[
\hat{\Theta}_{ab} = \begin{bmatrix}
a_0h_0 & a_1h_2 & a_2h_p & \\
a_1h_0 & a_2h_2 & a_3h_p & \\
\vdots & \vdots & \vdots & \\
a_{r-1}h_0 & a_0h_2 & a_1h_p & \\
a_1h_1 & a_2h_3 & a_3h_{p-1} & \\
\end{bmatrix}
\] (9)

The coefficients of the nonlinear and linear parts, $a=[a_0, \ldots, a_r]^T$, $b=[b_1, \ldots, b_p]^T$, can be resolved using Singular Value Decomposition (SVD) as:
\[
\hat{\Theta}_{ab} = U_1 \Sigma_1 V_1^T
\] (10)

The parameter vector estimates are obtained as follows:
\[
\hat{\theta} = \arg \min_{\theta} \parallel \hat{\Theta}_{ab} - \alpha \theta \parallel_2^2 = (U_1^T V_1)^{-1}
\] (11)
\[
\hat{a} = U_1
\] (12)
\[
\hat{b} = V_1 \Sigma_1
\] (13)

Bai proved that (12) and (13) are the best possible estimates for vectors $a$ and $b$. Bai also proved [8] that under rather mild conditions on the additive noise $w(n)$ and input signal $u(n)$ in (1), $\hat{a}(N) \rightarrow a$, $\hat{b}(N) \rightarrow b$, with probability 1 as $N \rightarrow \infty$. From equation (1) it can be seen that the coefficient sets $(b_k, a_i)$ and $(db_k, \alpha^t a_i)$ result in the same Hammerstein model. To obtain unique coefficients it is usually assumed that either $b_k$ or $a_i$ is to be normalized. It is assumed in (12) and (13) that $\|b\|_2 = 1$, i.e. the $a$-parameter vector is normalized.

The estimates of the linear coefficients are given by (13). If these linear coefficients do not yield a stable linear filter, the simple and well-known minimum phase correction is performed. Let $T_w$ represent the minimum phase transformation:
\[
\hat{b} = T_w(\hat{b})
\] (14)

Clearly, if $b \neq \hat{b}$, then the resulting Hammerstein model obtained by using $\hat{b}$ and $\hat{a}$ from (12) has no theoretical foundation; in other words $\parallel \hat{\Theta}_{ab} - \alpha \hat{\theta} \parallel_2$ may be very large. To compensate for the minimum phase correction we compute the $\hat{a}$ which minimizes $\parallel \hat{\Theta}_{ab} - \alpha \hat{\theta} \parallel_2$, in other words
\[
\hat{a} = \hat{\Theta}_{ab} \hat{b} (\hat{b}^T \hat{b})^{-1}
\] (15)

3. Minimum phase compensation test

The static nonlinear function for the test set was selected to be:
\[
N(a(n)) = \sum_{i=0}^{r} a_i u(n)^i
\] (16)

We used $r+1 = 3$ and two sets of $d$-vectors,
- Hammerstein model 1 (Ham1), $d_{0..2} = [0.33 0.5 1]$
- Hammerstein model 2 (Ham2), $d_{0..2} = [0.5 1 2]$

The parameters $d$ were selected based on our preliminary experiments. The linear filter order was selected to be 7. For comparison a $10^{th}$-order LPC-linear filter was also used in the tests. The parameter identification was accomplished using the method described in section 2. The well-known parameter regularization method called ‘Ridge Regression’, described in detail in [13], was implemented in the code. With this method the estimate of $\theta$ in (8) is changed to a form:
\[
\hat{\theta} = (\Phi \Phi^T + \alpha d^T d)^{-1} \Phi \hat{y}
\] (17)

Where I is the identity matrix and regularization parameter value $\alpha = 0.001$ was used in the test. In Fig. 1 an example of vowel signal and the filtered residuals are presented and in Fig. 2 an example of consonant signal /s/ and its residuals are presented. The scale above subfigures means that the scale factor is equal to maximum absolute value from signal in question. The signals were normalized between the values $[-1, 1]$ in order to show the signal information content rather than the amplitude response.

The test was made with 5 minutes of Finnish male and female utterances sampled at 8000 Hz. The experiment was done in Matlab environment. The frame size was 256 samples corresponding 32 ms of speech.

Under the original signal the $10^{th}$ order LPC filtered residual is presented. The compensated nonlinear and stable linear coefficient Hammerstein model processed signal is in the third subfigure and uncompensated/unstable Hammerstein model processed signals in subfigure 4. In Tables 1-4 are the linear and nonlinear coefficients for the Hammerstein model 1 (Ham1) used in the Figs. 1 and 2. The leading coefficient $b_0 = 1$ is left out from the linear coefficient tables.

| Table 1: Hammerstein linear filter coefficients for vowel /a/ |
|-----------------|-----------------|
|                | Unstable linear coefficients | Stable linear coefficients |
|                | -1.50 | 2.63 | -1.16 | 0.87 | 0.07 | -0.29 | 0.12 | -0.01 | -0.10 |
|                | -1.90 | -0.33 | 0.92 | -0.19 | -0.39 |

| Table 2: Hammerstein nonlinear coefficients for vowel /a/ |
|-----------------|-----------------|
|                | Uncompensated nonlinear coefficients | Compensated nonlinear coefficients |
|                | -0.42 | 0.90 | 0.14 |
|                | -0.94 | -0.31 | 0.15 |
Figure 1: Example male vowel /a/.

Table 3: Hammerstein linear filter coefficients for consonant /s/

| Unstable linear coefficients | 0.83 | -0.83 | -0.41 | -0.42 | 4.00 | -2.37 | 0.68 |
| Stable linear coefficients   | 0.39 | -0.34 | -0.44 | -0.13 | -0.14 | 0.28  | 0.17 |

Table 4: Hammerstein nonlinear coefficients for consonant /s/

| Uncompensated nonlinear coefficients | -0.06 | 0.47 | -0.88 |
| Compensated nonlinear coefficients  | 0.93  | 0.37 | -0.04 |

Figure 2: Example male consonant /s/.

The information content of the original and all residual signals were measured by Akaike’s Information Criterion (AIC). The results of the AIC test are in Table 5.

Table 5: The AIC test results for the original and residual signals

<table>
<thead>
<tr>
<th></th>
<th>Orig. LPC res.</th>
<th>St. Ham1</th>
<th>Ust. Ham1</th>
<th>St. Ham2</th>
<th>Ust. Ham2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC values for vowel /a/</td>
<td>-5.16</td>
<td>-5.19</td>
<td>-6.51</td>
<td>-2.59</td>
<td>-5.96</td>
</tr>
<tr>
<td>AIC values for consonant /s/</td>
<td>-7.29</td>
<td>-7.31</td>
<td>-15.60</td>
<td>-4.73</td>
<td>-11.46</td>
</tr>
</tbody>
</table>

In Table 5, “Orig” means the original signal information content, “LPC res.” is the LPC filtered residual information content, “St.” is abbreviation for stable and “Ust.” unstable Hammerstein models. The less information is in the signal, the more negative the result is. It seems that the balanced
Hammerstein model is removing the information better than the unbalanced Hammerstein model in both cases.

4. Conclusions
The Hammerstein model is fascinating approach for speech modeling purposes, and it can be viewed as an extension to linear predictive coding method. The parameter compensation method utilizes minimum phase correction for the linear coefficients. The influence of the minimum phase correction can be easily reflected to nonlinear function coefficients by using the original $\theta$ matrix in computations.

The stability in LPC models is kind of a ‘side-effect’ in parameter estimation. The stability of Hammerstein model depends on the linear coefficients, nonlinear functions (and coefficients) and their combinations. So, there are three different items and their stability to be ensured. Fortunately, the input-output relationship matrix defines the some of the conditions for parameter estimation and also for their stability.

Hammerstein model parameters can be solved, stabilized and balanced using method described in this paper. However the fundamental question about the type of nonlinearities, their modeling and practical implementation in speech processing still remains.

5. References