Subjective Evaluation of Nonlinear Signal Processing for Enhancing Perceptual Quality of Small Loudspeakers

Yusuke Nakashima, Nobuhiko Naka and Tomoyuki Ohya

Multimedia Laboratories
NTT DoCoMo, Inc.
nakasima@spg.yrp.nttdocomo.co.jp

Abstract
This paper shows the subjective test results for the performance of a preinverse filter compensating both linear and quadratic distortion of a small loudspeaker modeled by the 2nd order Volterra filter. The addition of the nonlinear preinverse filter improved the sound quality at least 0.5 in terms of Mean Opinion Score (MOS) for four of eight speech/music samples, and compensating quadratic distortion is found to be effective for six of the eight samples. Consequently, the results show the subjective effectiveness of nonlinear signal processing in improving the loudspeaker performance and thus mobile hands-free communication services.

1. Introduction
Loudspeaker playback is becoming a popular listening style in mobile communications. In addition to traditional speech telephony services, cellular phone users can enjoy various audio contents, such as video telephony, audio-visual clips, and rich customized ringing tones, on hands-free or hand-held style.

Codecs and loudspeakers play major roles in perceived audio quality for the above mentioned mobile applications. As a result of extensive research, state-of-the-art codecs now provide sufficient audio quality. For example, AMR[1] and AAC[2] defined in Packet Switched Streaming services[3] maintain quality close to the original, i.e., band-limited but not-encoded signal. The loudspeakers remain the weak link.

Loudspeaker performance attributes such as sound pressure level and bandwidth depend on the geometric parameters including diameter and allowable air space[4] which are in fact severe design constraints in mobile terminals. The constraints make it difficult to greatly improve loudspeaker performance. Moreover loudspeakers on mobile terminals are often driven at relatively high input level in order to ensure audibility in noisy environments. In this case, nonlinear distortion of the loudspeaker is noticeable. Additionally, enclosure, i.e., handset, vibration can degrade perceived sound quality.

Various models have been proposed for modeling loudspeaker nonlinearity. The Volterra filter based on Volterra expansion is a typical method to model the linear and nonlinear behavior of a loudspeaker[5]-[10]. The Volterra filter is basically not loudspeaker specific, and both linear and nonlinear response can be modeled at the same time. Compared to other existing methods, this generality is useful when applied to a loudspeaker system on a mobile terminal, but its complexity is huge. Therefore, complexity reduction is the main topic discussed in most papers that describe Volterra based loudspeaker linearizers. All performance assessments have used sinusoidal signals and been objective in nature. While these assessments show that Volterra-based linearizers are effective for the loudspeakers of mobile terminals[10], their improvement in quality remains unclear.

This paper shows subjective test results of Volterra-based mobile loudspeaker linearization. Subjective performance is measured in terms of MOS (Mean Opinion Score)[11] by using realistic signals, such as speeches and several kinds of music.

The next section explains the theory of the Volterra filter and its inverse filter. The third section describes the procedure used to construct a nonlinear preinverse filter that can compensate the linear and nonlinear distortions of a loudspeaker. The fourth section introduces our subjective test method. The fifth section describes the MOS difference noted in the subjective test. The sixth section concludes this paper.

2. Volterra filter
2.1. Representation of a Nonlinear System
The relation of the input $x(k)$ and output $y(k)$ of a 2nd order finite discrete system can be represented by the Volterra filter as[5]:

$$y(k) = \sum_{p=1}^{P} h_p [x(k)]$$

(1)

where the Volterra operator $h_p [x(k)]$ is described as,

$$h_p [x(k)] = \sum_{m_1=0}^{N_{p-1}} \cdots \sum_{m_p=0}^{N_{p-1}} h_p (m_1, \ldots, m_p) \times x(k-m_1) \cdots x(k-m_p),$$

(2)

where $h_p (m_1, \ldots, m_p)$ is called the $p$th order Volterra kernel, and $N_p$ is the memory length of the $p$th order
kernel which was set finite and equal to or more than zero.

The output of the 2nd order preinverse filter of (1) is depicted as [5]:

$$G_2(k) = \overline{g}_2[x(k)] - g_1[\overline{h}_1[x(k)]]$$

(3)

where \(\overline{g}_2\) is the inverse of \(\overline{h}_1\). Figure 1 shows the block diagram of the preinverse filter in (3).

![Figure 1: Second order preinverse filter of the second order Volterra filter.](image)

3. Modeling a nonlinear system

The response of a 15mm diameter loudspeaker was identified by a 2nd order Volterra filter. The specifications of the loudspeaker are described in Table 1. The loudspeaker was placed on a plate baffle, and a microphone was placed 50mm in front of the loudspeaker to measure its output; the frequency response is shown in Figure 2. The identification of the loudspeaker response by the Volterra kernels was performed by utilizing the linearity between the kernels and the output of the Volterra filter [6]. The tap length of \(h_1\) and \(h_2\) were set 400 taps and 4095 taps (90×91/2), respectively, at the sampling frequency of 65536 Hz. The symmetry of the kernels was utilized for reducing the kernel number of \(h_2\). The modeling accuracy of the response by the 2nd order Volterra filter was 31.5dB in signal to noise ratio (SNR) which was 5.4dB higher than the accuracy by the 1st order filter.

<table>
<thead>
<tr>
<th>Item</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td>15 mm</td>
</tr>
<tr>
<td>Rated input power</td>
<td>0.50 W (1.0 Wmax)</td>
</tr>
<tr>
<td>Impedance</td>
<td>8 ohm (2 kHz, 1 V)</td>
</tr>
<tr>
<td>Sound pressure level</td>
<td>98 dBSPL (Input: 0.5 W/50 mm at 1 kHz)</td>
</tr>
</tbody>
</table>

The nonlinear preinverse filter was then derived from the identified kernels based on (3). The linear inverse filter \(\overline{g}_1\) is produced by first deriving inverse filter of \(\overline{h}_1\) and convolving a band pass filter to suppress the unwanted amplification. The length of \(\overline{g}_1\) was 1688 taps. The compensation accuracy of the nonlinear preinverse filter, on a male speech for example, was 12.7dB in SNR which was 3.1dB higher than that of the linear preinverse filter.

4. Subjective evaluation test condition

The purpose of this subjective test is to evaluate the compensation performance of the nonlinear preinverse filter. Table 2 shows general conditions for the test.

The test methodology was Absolute Category Rating [11], where the subjects are asked to score the samples’ quality at one of 5 grades; 1: poor, 2: bad, 3: fair, 4: good, and 5: excellent. The test samples were presented to the subjects through headphones. Twenty-four non-expert subjects evaluated each sample twice with different randomization, thus 48 votes were obtained for each test sample in total.

For the test sources, the authors adopted two Japanese male speeches, two Japanese female speeches, and four types of music. Details of the music sources are described in Table 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test methodology</td>
<td></td>
<td>Absolute Category Rating (ACR)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1: poor, 2: bad, 3: fair, 4: good, 5: excellent</td>
</tr>
<tr>
<td>Source</td>
<td>8</td>
<td>4: Japanese speech</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: Music</td>
</tr>
<tr>
<td>Preinverse filter</td>
<td>3</td>
<td>Nonlinear preinverse filter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear preinverse filter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct</td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>1</td>
<td>2nd order Volterra filter model</td>
</tr>
<tr>
<td>Subject</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Randomization</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 2: Frequency response of the loudspeaker; solid line: linear response, broken line: 2nd harmonic, dotted line: 3rd harmonic.](image)
Table 3: Music sources of the subjective test.

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Length (seconds)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>14</td>
<td>Speech with BGM</td>
</tr>
<tr>
<td>Castanet</td>
<td>14</td>
<td>Castanet</td>
</tr>
<tr>
<td>Pop music</td>
<td>12</td>
<td>Several modern instrument</td>
</tr>
<tr>
<td>Piano</td>
<td>13</td>
<td>Piano</td>
</tr>
</tbody>
</table>

The sources were preinversed by each of the three preinverse filters as described in Figure 3; a) the nonlinear preinverse filter, b) the direct, and c) the linear preinverse filter. The nonlinear preinverse filter compensates the linear and the quadratic distortions of the loudspeaker. For the linear preinverse filter, \( \tilde{g} \), was adopted to compensate the linear distortion of the loudspeaker. The preinverse filter outputs were then fed to modeled loudspeaker, i.e., the 2nd order Volterra filter modeled in Section 3. The final outputs as well as sources were used as the test samples.

Figure 3: The procedure used to produce the test samples; a) nonlinear preinversed response, b) direct response, and c) linear preinversed response.

5. Results

The subjective test results for nonlinear preinversed responses were compared with those of following three types of samples; the direct responses, the linear preinversed responses, which are depicted in b) and c) in Figure 3, respectively, and the sources. For analysis, 95% confidence interval was used.

Figure 4 shows that for four of the eight sources, the nonlinear preinverse filter obtained 0.5 higher or more in MOS than the direct responses, which means the nonlinear preinverse filter was effective in compensating linear and nonlinear distortions.

Figure 5 shows that for six of the eight sources, the nonlinear preinverse filter obtained higher MOS than the linear preinversed responses, which implies that the nonlinear preinverse filter was efficient in compensating nonlinear distortions.

Figure 6 shows that for all the nonlinear preinversed responses other than castanet’s, the nonlinear preinversed responses were still worse in quality than the corresponding sources. The nonlinear preinversed response of News, for example, obtained higher MOS than the direct response and the linear preinversed response, but obtained lower MOS than the source by about 0.6. This reveals us that the nonlinear preinverse filter is still necessary to be improved its compensation performance.

Figure 4: MOS differences from the direct responses. Errorbars represent 95% confidence intervals (C.I.) of MOS differences. Gray bars indicate MOS differences are statistically equal or more than 0.5 in MOS with the C.I.

Figure 5: MOS differences from the linear preinversed responses. Errorbars represent 95% confidence intervals (C.I.) for testing the inexistence of MOS differences. Gray bars indicate MOS differences statistically exist with the C.I.
6. Conclusion

In this paper, we focused on distortions of small loudspeakers, and measured the subjective performance of the nonlinear preinverse filter to compensate the linear and quadratic distortions. The compensation performance of the nonlinear preinverse filter was evaluated using ACR. The test results show that the nonlinear preinverse filter was effective for half of the eight speech/music sources in improving the quality by compensating linear and nonlinear distortions. For more than half the sources, the nonlinear preinverse filter improved the quality by compensating the nonlinear distortions.

But there still exist the sources that the nonlinear preinverse filter was not perceptually effective in compensating the linear and quadratic simulated loudspeaker distortions. We noted that most nonlinear preinversed responses did not satisfactorily obtain the MOS of the sources with the use of the nonlinear preinverse filter, which is necessary to be compensated to achieve high quality in mobile hands-free services.

7. Acknowledgement

The authors wish to thank Dr. Khosrow Lashkari of DoCoMo Communications Laboratories USA, Inc. for his helpful suggestions.

8. References