Perceptual Study and Auditory Analysis on Digital Crossover Filters*

HENRI KORHOLA AND MATTI KARJALAINEN, AES Fellow
(hkorhola@gmail.com) (Matti.Karjalainen@tkk.fi)

Helsinki University of Technology, Department of Signal Processing and Acoustics, Espoo, Finland

Digital crossover filters offer interesting possibilities for sound reproduction, but there do not exist many publications on how they behave perceptually. In this research the phase and magnitude errors in digital implementations of linear-phase FIR as well as Linkwitz–Riley (L–R) crossover filters were studied perceptually and by auditory analysis. In headphone listening to crossover-processed signals the just noticeable level of degradation due to filter artifacts was explored. With a real loudspeaker experiment in a listening room the validity of the results was checked in realistic conditions and as guidelines for “safety limits” of filter orders that would not produce audible errors. For L–R crossovers the audibility of phase distortion can be predicted fairly well by group delay error. For high-order FIR crossovers this does not work because the perceptual mechanism is different: the errors become audible as ringing at the crossover frequency. This was explored by simple auditory modeling analysis, which can explain qualitatively the perceptibility of such artifacts.

0 INTRODUCTION

While digital technology has made its entrance into the audio world a couple of decades ago, crossover filters have remained mostly analog until very recent times. An interesting option for loudspeaker designers is a crossover filter with steep attenuation and constant group delay, approaching an “ideal” crossover filter [1]. This is possible with linear-phase FIR filters of sufficiently high order [2].

There are many publications on digital crossover filters [2]–[9], but hardly any perceptual studies. This would be important, however, because the perceived reality is often not coherent with signal processing theory. A lone exception is [10], which was published after this study had been launched. The two studies overlap slightly, though the approaches are different.

The present paper presents perceptual results from a study of linear-phase FIR crossover filters in different scenarios. Furthermore, digital implementations of well-known Linkwitz–Riley (L–R) crossover filters [11] were included in the experiment. Filter order, crossover frequency, and input signal were varied, as well as off-axis listening where misalignment of arriving signal components takes place.

Two listening experiments were conducted. The first was performed over headphones, and the two types of crossover filters were simulated in MATLAB [12]. In the second case an experiment with real loudspeaker and linear-phase FIR crossovers was conducted in a standardized [13] listening room. The goal of the loudspeaker experiment was to check practical guidelines for “safety limits” of FIR filter orders that do not produce audible errors.

The off-axis ringing phenomenon caused by a lack of time alignment [7], occurring with high-order FIR crossover filters, was of special interest in the listening tests. This is why the listening angle was varied in addition to the filter properties.

Following this introduction, the paper continues in Section 1 with background knowledge of crossover filters and a brief review of related psychoacoustics. The listening experiments are described in Section 2, the main results in Section 3, and a more in-depth analysis is given in Section 4. Auditory modeling analysis is presented in Section 5, followed by conclusions.

1 BACKGROUND

Crossover filters are special filters that divide the audio spectrum for loudspeaker drivers so that the drivers can operate within their optimal frequency ranges. Crossover filters are a necessity in multiway loudspeakers. A basic division of crossover filters can be passive, active, and digital crossovers.
In a two-driver case the transfer function of a crossover filter consists of low- and high-pass outputs,
\[ H(j\omega) = H_L(j\omega) + H_H(j\omega) \]  
(1)
where \( H_L(j\omega) \) and \( H_H(j\omega) \) are the transfer functions of the low- and high-pass subfilters, respectively, and \( \omega = 2\pi f, \) \( f \) being the frequency.

Crossover filters should protect the drivers from unwanted frequency components, which means that their attenuation capabilities should be adequate, but at the same time they should affect the reproduction as little as possible. The goals of crossover design (in decreasing order of importance) can be summarized as follows, adopting from Lipshitz and Vanderkooy [1] and Linkwitz [11]:

1) Flatness in magnitude response of the combined low- and high-pass outputs on the main listening axis; \( |H(j\omega)| \approx \text{constant} \).
2) Steep enough band-stop slopes of the subfilters to limit the reproduced frequency range for each driver.
3) Acceptable polar response, that is, phase difference between low- and high-pass outputs approximately zero,
\[ \phi_L(\omega) - \phi_H(\omega) = \arg[H_L(j\omega)] - \arg[H_H(j\omega)] \approx 0 \]  
(2)
4) Acceptable phase response being linear at best,
\[ \phi(\omega) = \arg[H_L(j\omega) + H_H(j\omega)] \approx k\omega. \]  
(3)
The group delay is a commonly used measure of phase distortion in crossover filter analysis. It indicates how much a certain frequency component or frequency range of a signal is delayed (the first derivative of phase),
\[ \tau_g(\omega) = -\frac{d\phi(\omega)}{d\omega}. \]  
(4)
It is also called envelope delay as it tells how much the envelope curve of a complex signal that contains many frequencies is delayed. It is usually given in either milliseconds or samples.

1.1 Crossover Filter Types

Passive crossovers have been the majority in the past, but active crossover filters have become increasingly common, especially in professional audio. Digital crossover filters are newcomers in the audio field.

1.1.1 Linkwitz–Riley Crossover Filter

The L–R crossover filter is one of the best known active crossovers [11]. It consists of low- and high-pass subfilters, each being a cascade of two identical Butterworth stages. For the simplest case with second-order subfilters and ideal drivers the total transfer function is
\[ H(s) = H_L(s) - H_H(s) = \frac{1}{(1 + s_n)^2} \times \frac{s_n^2}{(1 + s_n)^2} = \frac{1 - s_n}{1 + s_n} \]  
(5)
where \( s_n \) is the Laplace transform variable normalized to the crossover frequency \( f \), that is, \( s_n = s/2\pi f \). The result is a first-order all-pass filter.

The L–R crossover has a flat magnitude response on the main listening axis and relatively good off-axis response. The commonly used fourth-order L–R crossover has an attenuation rate of 24 dB per octave, which is adequate in most cases. The combined phase response is nonlinear, producing phase distortion to the output signal, and so the question is: when is it audible? Linkwitz himself concluded in the original paper from 1976 as well as on his Web pages that at least with a fourth-order L–R filter, the phase errors are not audible. We also tried to find guidelines for the audibility of L–R phase errors in our headphone-based listening experiment (Section 2). It is already known that the audibility of phase distortion is individual and depends on the sound material used [14].

1.1.2 FIR Crossover Filter

FIR crossover filters can offer so-called brick-wall filtering, which means a very steep separation of the frequency spectrum. Another interesting property is the possibility of linear-phase reproduction. Importantly, it happens only when the low- and high-pass outputs are perfectly summed for the combined output, that is, on the listening axis, when the time delay between drivers is zero. The transfer function of an FIR filter is
\[ H(z) = \sum_{n=0}^{N} h(n)z^{-n} \]  
(6)
where \( N \) is the order of the filter, \( h(n) \) is the discrete-time impulse response, and \( z \) is a complex variable in the \( z \)-transform domain.

1.2 Psychoacoustics

There can be two reasons for audible errors. As far as magnitude errors are concerned, a general rule of just noticeable differences (JND) is about a 1-dB deviation if the sound pressure level is moderate and there is a comparable reference in the short-term memory. While the general opinion says that the magnitude response plays a major role in sound perception, the phase response should not be neglected. Many studies have been published regarding the perception of phase distortion [14–24], and the conclusions are the following.

1) Phase distortion is audible with certain signals. Impulsive sounds are most susceptible.
2) Perception of phase distortion is individual; there are clear differences between subjects.
3) Phase distortion is generally more audible with headphones than with loudspeakers.
4) A rule of thumb for the audibility limit of group delay deviations would be about 1.6 ms, independent of

Notice that the lengths of coefficient and state variable vectors for an FIR filter are \( N + 1 \). For an IIR filter the order is the maximum of the numerator and denominator polynomial orders.
the center frequency, when studied with all-pass filters, according to [24].

In our experiments we searched for guidelines of just perceptible phase errors in crossover filters with different signals. We will discuss group delay errors, which means here the peak-to-peak variations of group delay over the crossover frequency range.

As will be found, the magnitude and group delay errors are not enough to explain the sound degradations in FIR crossover filters, and simplified auditory modeling is presented in Section 5 in order to explain the related perceptual phenomena qualitatively.

2 LISTENING EXPERIMENTS

The two crossover filters under study are the linear-phase FIR crossover and digital implementation of the L–R crossover. The more interesting part of the experiments was the linear-phase FIR crossover, because we wanted to find out how its brick-wall filtering with linearity in the phase response affects sound material with different parameters perceptually.

The goals of the study were set as follows:

1) How much deviation from uniform group delay is allowed for different signals, that is, what is the JND limit? How does the type of group delay error affect audibility?
2) How sensitive are digital crossover filters for off-axis response errors?
3) Is it possible to predict the results of the listening test by objective measures?
4) Which approximate “safety limits,” in particular for FIR filter orders, would not produce audible errors?

The experiments were conducted as comparisons between the original, unprocessed sound sample and the crossover-processed sound sample. Headphone listening of the crossover-processed signals (test 1) was chosen as the primary test method to reveal small degradations. This simulates listening to a loudspeaker with ideal drivers in an anechoic room. A smaller scale listening test was conducted with a real loudspeaker in a listening room (test 2) to check the validity of the results from the first test under realistic conditions.

2.1 Test 1: Crossover Listening by Headphones

In addition to the first author, nine subjects with some experience in listening tests listened to the comparison samples through headphones in random order, using a graphical user interface (GUI), implemented in MATLAB. Audiometry tests were not performed, but the subjects did not have known hearing degradations. They were supposed to give grades for the basic audio quality difference between unprocessed and crossover-filtered samples. An oral introduction to the topic, instructions, and guidance were provided prior to the test.

The grading was based on Table 1 of ITU’s small impairment scale [13], which is used for judging small distortions in audio signals. The scale is guided to be used at intervals of 0.1, but based on feedback from preliminary tests, it was used at intervals of 0.5 instead. Comprehensive training would have been necessary for using steps of 0.1. The only trained subject for this test was the first author because of preliminary listening in different scenarios. As the samples were graded in randomized order using the GUI, the first author could take part in the test.

### 2.1.1 Test Material and Parameters

The test material consisted of three different sound samples:

1) 10-Hz square wave
2) Castanets
3) Tom-tom drum.

The selection of test material was based on preliminary tests as well as on general knowledge of psychoacoustics. The square wave is known to be highly revealing for phase distortions in audio systems. It has been used in many perceptual studies, such as [14], [20]. The castanets were chosen as a test sound because of their transient nature. The castanet percussions are concentrated in a short period of time, and so the distortions would affect the final waveform in an audible way. The third sample was the tom-tom drum. It also has a slight impulse nature in its waveform, but not as concentrated as the castanets. It was considered an interesting sample, because drums are very common musical instruments, and in contrast to the castanets, they have low-frequency contents.

Defining the parameters for the crossover filters was planned carefully. The crossover frequency was chosen to be at either 100/300 Hz, 1 kHz, or 3 kHz to approximate commonly used divisions in multiway loudspeakers. The orders of the crossover filters were chosen to be realistic for practical implementations, and not to exceed 32 for L–R or 2000 for FIR crossovers. Some exceptions were included. The delay between low- and high-pass outputs was implemented to simulate the elevation of the listening angle, because most of the problems exist when the vertical angle is changed. According to a quite normal separation of drivers in a two-way loudspeaker (0.25 m), the delay was limited to a range of 0–0.5 ms to simulate far-field elevation angles between 0 and 45°.

### Table 1. ITU small impairment scale.*

<table>
<thead>
<tr>
<th>Impairment</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperceptible</td>
<td>5.0</td>
</tr>
<tr>
<td>Perceptible, not annoying</td>
<td>4.0</td>
</tr>
<tr>
<td>Slightly annoying</td>
<td>3.0</td>
</tr>
<tr>
<td>Annoying</td>
<td>2.0</td>
</tr>
<tr>
<td>Very annoying</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*Scale was used in listening test at intervals of 0.5.
The parameters are listed in Table 2, illustrating the diversity of possible scenarios. The large number of different scenarios forced us to cut out inaudible samples, making the test reasonable in size and duration. For the same reason, the parameters were not always the same for different signals.

### 2.1.2 Test Equipment

The listening test was carried out with a laptop computer, which was equipped with MATLAB-based software for user interface. To ensure sound quality at an adequate level, an external sound card Fast Track Pro of M-Audio was used. The headphones used in the experiment were Sennheiser HD-600 headphones.

### 2.2 Test 2: Crossover Listening with a Real Loudspeaker in a Room

A real loudspeaker was used for testing to find out the differences between headphone simulation and real listening in a room, and also to determine approximate “safety limits” for FIR crossover filter orders under realistic listening conditions. In this case the crossover filter under study was a linear-phase FIR crossover at the crossover frequency of 3 kHz only. The orders of low- and high-pass subfilters varied from 300 to 2400. The loudspeaker was a two-way loudspeaker with a bass-reflex enclosure. The enclosure’s volume was 12 liters. The loudspeaker was equipped with bass-range driver L18RNX/P and tweeter 25TAFN/G of Seas [25]. The magnitude responses of the woofer and tweeter are illustrated in Fig. 1. As the crossover artifacts are probably occurring only in a very narrow band, equalization of the responses was considered unnecessary. The woofer was 0.78 m and the tweeter 0.92 m above the floor. The room was a standardized [13] listening room with a reverberation time of about 0.3 s at 500 Hz [26].

### 2.2.1 Test Procedure

Five untrained but audio-related persons participated in the experiment. After analyzing the results from headphone listening, it was decided that the test would serve to find the FIR filter orders that do not produce audible errors. The test signals used in the loudspeaker experiment were the 10-Hz square wave and the castanets. The experiment consisted of five different listening positions. In each the subject listened to the signals both sitting on a chair and standing. The positions in the listening room are depicted in Fig. 2. The loudspeaker was placed as the left channel in stereo listening (for position 1). Position 5 was directly in front of the loudspeaker so that the subject was directed toward the loudspeaker, which was different from positions 1 to 4.

The threshold for just audible errors was searched for by playing pairs that consisted of the reference signal of low-order (300) FIR crossover and the test signal with higher order (600–2400) FIR crossover. Based on preliminary tests it was assumed that the 300th-order filter artifacts were not audible. The sample pairs were played, varying the filter order up and down until an approximate value of just noticeable degradation was found.

### 3 RESULTS FROM LISTENING EXPERIMENTS

In this section an overview of the results from the listening experiments is given. A more in-depth analysis and a discussion are presented in Section 4.

#### 3.1 Results of Headphone Listening Experiment

Since there were so many varying factors, as shown in Table 2, presenting and interpreting the results is not straightforward. The total count of different samples is 79,
which are divided into 27 subgroups. In each subgroup some parameter is varied, and its effect on the perception of distortions is studied. It must be emphasized that based on the preliminary tests, in many of the possible scenarios distortion was considered inaudible, and they were left out of the test.

The chosen combinations of results are presented for both FIR and L–R crossovers. Different subgroups have been plotted in the same plot in order to illustrate the differences in perception between signals. The results are plotted with MATLAB’s boxplot function. It plots the median as the dashed horizontal line inside the box. Lower and upper quartiles are plotted as the bottom and top lines of the box. Whiskers, the dashed lines, extend from the ends of the boxes to the adjacent values of the data up to a maximum of 1.5 times the interquartile range. The grades not in the range of 1.5 times the interquartile are plotted as outliers by an $\times$.

### 3.1.1 Results for FIR Crossover Filters

The grading results for 1000th- to 2000th-order FIR crossovers at 1 kHz are plotted in Fig. 3. It can be noticed that a 0.2-ms delay between drivers causes clearly perceptible distortions in the square wave. With a 2000th-order FIR at 1 kHz, castanets are susceptible to distortions rather easily, so that a 0.3-ms delay between the low- and high-pass bands produces a perceptible distortion. The tom-tom drum is the most insusceptible of the signals used; it does not suffer from distortion with even a 0.5-ms delay at 1 kHz.

Results for the two other crossover frequencies (100 Hz and 3 kHz) are plotted in Fig. 4. Excluding a couple of outlying gradings, the results for FIR crossovers at 100 Hz show that even very high orders do not produce audible distortions in any of the signals with the maximum delay between drivers (0.5 ms). This is illustrated in Fig. 4(a).

On the contrary, differences become clearly audible at 3 kHz. The results for an FIR crossover at 3 kHz with a maximum delay of 0.5 ms are presented in Fig. 4(b). As the order of the crossover filter is varied, it seems that castanets begin to suffer from distortions somewhere between orders 700 and 1000. The tom-tom presents clearly audible distortions at order 700.

When testing the highest order of 2000 with FIR crossovers, interesting results occur, as Fig. 4(c) shows. After the smallest delay of 0.01 ms, all signals start to present audible distortions. The delay is really small, but still the medians of the results are clearly in the disturbance range (below or at grade 4). These results are interesting because this type of FIR crossover should offer the “ideal” frequency-domain characteristics of a crossover with very steep stopband attenuation, yet linearity in the phase response. However, while the audio spectrum is divided steeply into low- and high-pass bands, ringing in the time domain occurs because of the Gibbs phenomenon [27]. This will be discussed in Section 4.4.

### 3.1.2 Results for L–R Crossover Filters

The digital implementation of L–R crossovers was used to study the audibility of phase distortion. The L–R cross-
over offers a flat magnitude response on the main axis, but it suffers from phase distortion due to increasing group delay deviations as the order increases. Combined results of the L–R cases are presented in Fig. 5. When the crossover frequency is at 300 Hz, the square wave presents more audible errors than the castanets or the tom-tom, as shown in Fig. 5(a). Looking at the L–R crossover of order 16, the square wave results received a median grade below 2, whereas the castanets and the tom-tom have median grades above 4.

At 1 kHz the square wave once again suffers from distortion above order 8, whereas the castanets received acceptable median of grading up to order 16 and the tom-tom does not seem to present considerable distortion up to order 24. The results of the L–R listening test at 1 kHz for all three signals are plotted in Fig. 5(b).

With the crossover frequency at 3 kHz, higher orders than with lower crossover frequencies seem to be acceptable in light of the results. With the square wave, for orders up to 20, and with castanets and tom-tom up to 32, the median grades are clearly above 4 [Fig. 5(c)].

### 3.2 Results of Loudspeaker Listening Experiment

In the listening experiment with a real loudspeaker and FIR crossover at 3 kHz in a listening room the degradations were inaudible to all test subjects at order 600 for the 10-Hz square wave and at order 900 for the castanets. Above these orders most of the test subjects perceived clear degradations. Because of the high sensitivity of the results to various factors, especially due to minimal changes in the listener head position, we did not collect extensive statistics of the results but rather tried to find only the lowest filter order where any artifacts appeared.

The “safe” order that does not produce perceived errors seems to follow that obtained in the headphone listening experiment for the 10-Hz square wave (see Fig. 10), but was larger for the castanets. This is a natural occurrence, especially for real-life signals, since the reflections in a room make the perception of phase more difficult as compared to headphone listening without reflections.

The lowest orders that produced noticeable errors were found in front of the loudspeaker in position 5 (see Fig. 2),

---

**Fig. 4.** Combined grading results of listening test for FIR crossovers at 100 and 300 Hz. (a) 0.5-ms delay, all signals. At 100 Hz no considerable degradations are perceived. (b) 0.5-ms delay, castanets and tom-tom. At 3 kHz with long delay, grades seem to decrease after order 600–700. (c) 2000th-order crossover, all signals. At 3 kHz with varying delay, even small delays seem to degrade signals in an audible way.
when the subject was sitting on a chair. This is likely due to the dominance of direct sound over reflections and to the directivity of the loudspeaker. Qualitative comments from the subjects suggested that the ringing of FIR crossovers was greatly dependent on the listening place. Slight changes in a subject’s head position could make the phenomenon either audible or inaudible.

The main finding is that the errors were clearly audible with a real signal and a real loudspeaker in a listening room for orders above 600 with the 10-Hz square wave, which is well in line with the results from the headphone listening test. This suggests that the headphone listening results can also be used as guidelines for practical listening conditions to avoid artifacts from crossovers.

4 ANALYSIS AND DISCUSSION

The analysis of the listening results concentrates on the computable measures of magnitude error and group delay error. In order to make the magnitude errors interpretable in terms of psychoacoustics, a one-third-octave smoothed spectrum was calculated in MATLAB. It is a rough measure of spectral resolution related to the critical bands of hearing. L–R crossovers were tested both on and off axis, whereas FIR crossovers were tested only off axis because the on-axis response should be nearly perfect.

Group delay deviations (peak to peak) were used as a measure of phase errors.

4.1 Magnitude Errors of L–R Crossover Filters

For L–R crossovers having a good off-axis response, and for FIR crossovers approaching ideal cutoff with steep attenuation, one-third-octave spectra exhibited hardly any audible magnitude errors among the 79 test samples. The criterion of audibility limit of spectral deviation is kept at 1 dB. The few cases with a larger magnitude error and a low average grade from the listening test are not so interesting from the point of view of this study. They were square wave signals filtered with 4th- and 16th-order L–R crossovers and having a 0.2-ms delay between drivers, suffering from 8.8-dB and 3.9-dB peak magnitude errors. Regardless of that, the average grades for them were as

![Fig. 5. Combined grading results of listening test for L–R crossovers, all signals. (a) At 300 Hz only square wave seems to be degraded audibly while castanets and tom-tom have quite high grades. (b) At 1 kHz square wave suffers from degradation above order 8, castanets above order 16, and tom-tom does not present clear distortion, even at order 24. (c) At 3 kHz degradations become audible with square wave above order 20, but castanets and tom-tom do not suffer from considerable distortions up to order 32.](image-url)
good as 4.0 and 4.1, which suggests only a minor perceptible change in the signal.

4.2 Group Delay Errors of L–R Crossover Filters

The JND limits of group delay distortions can be approximated from the results. They are shown in Table 3. Care must be taken when interpreting the results, but they offer guidelines for group delay limits. As a general conclusion, phase errors larger than roughly 1.5–2 ms due to L–R crossover filtering may become audible, which is in accordance with [23].

Plotting the results from a different point of view may give a better insight into the problem. Grades from the listening test as a function of group delay are plotted in Figs. 6–8. Looking at the figures, the grades and group delay errors seem to have an interdependence. According to Figs. 7 and 8 the real-world signals, the castanets, and the tom-tom would accept group delay errors of 3 ms or more before they become audibly distorted. No odd behavior is found in the correlation between grades and group delay errors, so these rough guidelines can be for the phase distortion. Hence point 1) of the experiment goals (Section 2) is answered for L–R crossovers, though no exact JND limits for different signals can be stated.

4.3 Group Delay Errors of FIR Crossover Filters

The JND limits of group delay errors for FIR crossovers are not so straightforward, as we will find out. The grades from the listening test as a function of group delay error for the square wave are plotted in Fig. 9. Notice that here the group delay error grows due to an increase in the time delay between low- and high-pass bands. No strange behavior exists, and the slopes descend almost similarly as for the L–R crossovers in Fig. 6.

The grades from the listening test as a function of group delay errors for the castanets are plotted in Fig. 10.

Table 3. JND limits for audible group delay errors with L–R crossovers.*

<table>
<thead>
<tr>
<th>Signal</th>
<th>Crossover [Hz]</th>
<th>Group Delay JND Limit [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square wave</td>
<td>300</td>
<td>1.8–4</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.25–4.9</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>1.3–2.4</td>
</tr>
<tr>
<td>Castanets</td>
<td>300</td>
<td>Over 9.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>3–4.9</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>Over 2.4</td>
</tr>
<tr>
<td>Tom-tom</td>
<td>300</td>
<td>Over 9.8</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>Over 4.9</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>Over 2.4</td>
</tr>
</tbody>
</table>

Due to sparsity of data, exact limits cannot be concluded.

Fig. 6. Grade as a function of group delay error for 10-Hz square wave signal with L–R crossovers. There is no delay between drivers. Graphs behave regularly, descending as group delay error increases.

Fig. 7. Grade as a function of group delay error for castanets with L–R crossovers. There is no delay between drivers. As with other signals, slopes are descending, though not as abruptly.

Fig. 8. Grade as a function of group delay error for tom-tom with L–R crossovers. There is no delay between drivers. As with other signals, slopes are descending, but rate is even smaller.

Fig. 9. Grade as a function of group delay error for 10-Hz square wave with FIR crossovers. Delay between drivers is varied. Graphs behave regularly, descending as group delay error increases.
Remembering that the castanets and the tom-tom were quite insusceptible to group delay errors, interesting results appear. The series of FIR crossovers with varying order and fixed delay between low- and high-pass channels fits badly into the picture. The other series in Fig. 10 show a nice behavior as a function of group delay, but the last series has to be extracted to a plot of grade as a function of the filter order [Fig. 10(b)]. It reveals that beyond order 700 the grade begins to decrease dramatically.

Examining Fig. 11 we see the same kind of behavior for the tom-tom as for the castanets. The series of FIR crossovers with varying order does not behave at all like the other series, which suggests that the value of the group delay error is not explaining the perceived degradation. It is again extracted to Fig. 11(b). The graph suggests that the tom-tom is even more susceptible to errors with increasing order than the castanets. Immediately after order 500 the problems begin, and the audio quality is no longer acceptable.

The results imply that in case of very small magnitude errors the group delay deviations can explain the decrease in audio quality to some extent, but as Figs. 10 and 11 clearly show, predicting audible errors by the values of the group delay errors is not always possible. Hence point 1) of the experiment goals (Section 2) so far remains unanswered for the FIR crossover case.

4.4 Ringing Phenomenon in FIR Crossover Filters

An explanation for the weird behavior of the grade versus group delay plots is the ringing phenomenon, which occurs with FIR filters as the filter order increases. It happens because of the Gibbs phenomenon, which is the aftermath of the very steep attenuation of the low- and high-pass filters [27]. On axis, with ideal drivers, it is no problem because the low- and high-pass impulse responses sum up nicely, and the crossover filter impulse response is just a delayed impulse. Off axis the summing does not succeed and residual ringing will remain.

A comparison is made between two FIR samples with castanets. Table 4 lists the parameters. According to the group delay error values it would be realistic to expect a better grade for a 2000th-order FIR crossover. However, the average grade of the 700th-order FIR is 4.4, whereas the 2000th-order FIR has only received an average grade of 2.3. This seems counterintuitive to judge by the values of the magnitude and group delay errors. The only factor that explains the difference is the much higher order of the crossover filter, apparently offering impressive properties for crossover filtering, but eventually degrading the signal substantially.

With regard to the ringing in the time domain, Fig. 12 presents zoomed plots of the impulse responses of 700th-
and 2000th-order FIR crossover filters computed for an off-axis position. As the theory [27] dictates, the height of the ripple is unchanged when the filter order is increased, but the time span of the ripples grows. This is seen clearly in Fig. 12. For order 2000 the ringing lasts longer in the time domain on both sides of the main response.

A temporal masking in hearing is the reason for the result. Impulsive sounds mask quite symmetrically in time [28], so both pre- and postringing may become audible, especially when we notice that with linear-phase FIRs they exist because of the symmetry of the impulse responses. Often, with real-life signals, the signal itself masks the error. Because of the sharp rise and decay of the castanets and the sharp rise of the tom-tom signals, audible errors appear easily. As seen in Fig. 4, the castanets show susceptibility to errors at higher frequencies (1 and 3 kHz). With the tom-tom the attack (rise) of the waveform is even steeper than for the castanets, which could explain its susceptibility to ringing phenomenon errors. The decay of the tom-tom waveform is slow compared to the percussion of the castanets, but preringing could explain the errors.

Another comparison of the ringing phenomena of the L–R filtered and FIR filtered tom-tom drum sample is given in Table 5, which lists the parameters of the samples, group delay error values, magnitude error values, and average grades obtained.

As can be noticed, group delay errors are quite modest for the FIR crossover (only 2.4 ms for the L–R and 1.57 ms for the FIR). The magnitude error is zero for the L–R because of listening on axis, and it is only 0.01 dB for the FIR according to the smoothed one-third-octave spectrum. Thus one could expect that the samples would receive approximately the same grades. However, the average grades are far apart, as the L–R crossover has received a grade of 4.9 and the FIR crossover only a grade of 3.1. The audible error that causes the low grade for the FIR is a kind of “squeak” in the beginning of the drum hit, which is caused by the preringing phenomenon in the impulse response.

To study the problem more deeply, the impulse responses are examined first. They are plotted in Fig. 13. There exists some ringing in the L–R impulse response, but it practically dies off after 5–6 ms. The duration of the FIR ringing in the impulse response is much longer. Worth noticing is also the asymmetry of the L–R impulse response. As the FIR crossover is realized with linear phase, the impulse response has to be symmetric [27]. This makes the preringing phenomenon more easily audible, as the signal itself masks the errors well forward with fast rising but slowly decaying sounds, such as the tom-tom. The same does not happen backward.

Fig. 14 shows the behavior of the group delay of these crossover filters. Although the L–R crossover group delay graph [Fig. 14(a)] is not smooth, it has no abnormalities in it. The FIR crossover group delay graph [Fig. 14(b)] has a very sudden change around the crossover frequency.

<table>
<thead>
<tr>
<th>Order</th>
<th>Cross over [Hz]</th>
<th>Delay [ms]</th>
<th>Signal</th>
<th>Group Delay Error [ms]</th>
<th>Magnitude Error [dB]</th>
<th>Average Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>3000</td>
<td>0.5</td>
<td>Cast</td>
<td>2.25</td>
<td>0.6</td>
<td>4.4</td>
</tr>
<tr>
<td>2000</td>
<td>3000</td>
<td>0.5</td>
<td>Cast</td>
<td>0.93</td>
<td>0.2</td>
<td>2.3</td>
</tr>
</tbody>
</table>

*Delay between drivers—0.5 ms. Magnitude error is decreasing from 0.6 to 0.2 dB. Judging from group delay and magnitude error values, 2000th-order FIR should receive much better grade, but opposite is observed.

Fig. 12. Comparison of impulse responses of 700th-order and 2000th-order FIR crossovers at 3 kHz in an off-axis position (delay between drivers is 0.5 ms) to demonstrate ringing phenomenon. Notice how ringing duration grows with increasing filter order. (a) 700th-order crossover. (b) 2000th-order crossover.
Table 5. Comparison of tom-tom samples demonstrating in ringing phenomenon.*

<table>
<thead>
<tr>
<th>Type</th>
<th>Order</th>
<th>Cross over [Hz]</th>
<th>Delay [ms]</th>
<th>Group Delay Error [ms]</th>
<th>Magnitude Error [dB]</th>
<th>Average Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-R</td>
<td>32</td>
<td>3000</td>
<td>0</td>
<td>2.4</td>
<td>0</td>
<td>4.9</td>
</tr>
<tr>
<td>FIR</td>
<td>2000</td>
<td>3000</td>
<td>0.03</td>
<td>1.57</td>
<td>0.01</td>
<td>3.1</td>
</tr>
</tbody>
</table>

*Delay implemented to FIR crossover is very small. Group delay errors are below audibility limits received. Magnitude error of FIR is increasingly small, only 0.01 dB, which is well below the general perception level of 1 dB. Nevertheless, average grades differ remarkably.

Fig. 13. Comparison of impulse responses of 32th-order L-R and 2000th-order FIR crossovers at 3 kHz to demonstrate ringing phenomenon. Notice asymmetry in ringing of L–R’s versus symmetry of FIRs. Impulse responses are closely zoomed to illustrate phenomenon better. (a) L–R crossover, order 32. (b) FIR crossover, order 2000.

Fig. 14. Group delays of L–R and FIR crossovers, at 3 kHz, 0.03-ms delay. Notice abrupt changes of FIR crossover group delay around crossover frequency in zoomed subplot. (a) L–R crossover, order 32. (b) FIR crossover, order 2000.
The essential question is: which one is audible, the minimal ripple in the magnitude response or the strange behavior of the group delay response? Both degradations are below the generally known and obtained audibility limits of group delay (see the L–R results in Table 3) and of magnitude. An explanation for errors with FIR crossovers should be explored with time-domain analysis.

However, point 2) of the experiment goals (Section 2) can be answered after these two inspections: high-order FIR crossovers seem to be highly susceptible to off-axis errors. The delay of 0.03 ms corresponds only to a flight-time difference of 10 mm between the loudspeaker drivers. This equals an off-axis position of only 2–3° when the loudspeaker drivers are separated by 0.25 m.

4.5 Conclusions from Listening Experiments
1) For L–R crossovers the JND limits for group delay deviations vary between signals, the 10-Hz square wave being the most susceptible, the castanets being less susceptible, and the tom-tom being the least susceptible to phase distortions. The guidelines for JND limits can be obtained from Table 3.
2) For FIR crossovers the JND limits for group delay errors cannot be obtained as no systematic correlation between group delay error values and grades exists.
3) The shape of the group delay graph might be crucial for perceived errors. FIR crossover group delay curves show irregularities compared to L–R crossover group delay curves.
4) FIR crossovers seem to be highly susceptible to off-axis errors with higher filter orders. A flight-time difference of only 0.02–0.03 ms between low- and high-pass bands at 3 kHz was found to produce audible ringing with high FIR orders of 1000–2000. Rough safety limits would be to keep the order of a linear-phase FIR filter at or below 600 at 3 kHz, according to the headphone simulation and the real loudspeaker experiment.
5) Predicting the perceptual results for L–R crossovers can be done, to some extent, by the group delay error values. Prediction for FIR crossovers seems to be more complicated, as the ringing caused by the Gibbs phenomenon causes different behaviors with higher filter orders.
6) The apparent characteristics of an “ideal” crossover filter turn out not to be pursued at any cost, because brick-wall attenuation with a linear-phase response demands a high-order FIR filter, which is exposed to the ringing phenomenon.
7) It seems obvious that the perceptual effect of ringing has to be studied in the time domain by some kind of auditory modeling, as will be done in the next section.

5 AUDITORY MODELING ANALYSIS

5.1 Structure of Filter-Bank Model
A simple auditory modeling analysis was used to interpret the results of the listening tests. Because the degradations in the signals caused by the crossover filters occurred in a narrow band, the external and middle ear modeling were left out, assuming that the auditory response is relatively flat in the crossover region. Furthermore, in the auditory filter-bank analysis only the critical bands in the crossover region were applied. The auditory analysis, implemented in MATLAB, was made maximally straightforward, and it consisted of the following steps:

1) Zero-phase (forward and backward) bandpass filtering using a bandpass Butterworth filter with a bandwidth corresponding to the Bark scale [29],

\[ \Delta f_{cb}[\text{Hz}] = 25 + 75 [1 + 1.4(c_c[\text{kHz}])^{2}]^{0.69} \]  (7)

where \( f_c \) is the characteristic frequency of the auditory channel.

2) Full-wave rectification (that is, absolute value) instead of half-wave hair-cell rectification to smooth the envelope response.

3) Monaural time resolution by 3rd-order Butterworth low-pass filtering at 300 Hz for simulating auditory nerve synchrony. No adaptation or temporal integration was used.

Zero-phase filtering with Butterworth-type bandpass filter was done with MATLAB’s filtfilt function. A second-order bandpass filter was applied three times in both forward and reverse directions, and a zero-phase filter was thus obtained with smooth, symmetrical responses. Using a single higher order bandpass filter without reverse filtering (that is, nonlinear phase) was found to produce more ripple in the auditory channel envelope response, which made the interpretation more difficult; thus it was omitted. The time alignment of original and crossover-processed signals was also easier when using a zero-phase bandpass filter.

The magnitude responses of the filter bank used in three auditory channels are illustrated in Fig. 15. The main auditory channel’s center frequency corresponds to the crossover frequency, which in this case is 3 kHz.

![Fig. 15. Magnitude responses of filter bank’s three channels.](image)

J. Audio Eng. Soc., Vol. 57, No. 6, 2009 June
Lower (center frequency 2.75 kHz) and upper (center frequency 3.25 kHz) adjacent channels are also plotted in Fig. 15. The −6-dB bandwidth is approximately 260 Hz.

5.2 Auditory Responses to 10-Hz Square Wave

Auditory modeling analysis was applied to crossover processed (FIR and L–R) signals (10-Hz square wave and castanets) to visualize the auditory responses in the time domain.

5.2.1 FIR Crossover Response

Both the rising and the falling edges of the 10-Hz square wave signal cause an auditory excitation in critical band filters. The auditory responses on three channels (crossover channel at 3 kHz, lower adjacent channel at 2.75 kHz, upper adjacent at 3.25 kHz) for a 2000th-order FIR crossover filter are shown in Fig. 16. The responses to the original signal are symmetrical and emerging at constant intervals of 50 ms, as expected from the square waves transitions.

The spreading of the time-domain response for the FIR 2000th-order crossover at 3 kHz is clear when the delay between the loudspeaker drivers is 0.02 ms. Though behaving quite nicely on the lower and upper auditory channels, the crossover-processed signal response is vastly spread on the auditory channel of 3 kHz, which corresponds to the crossover frequency. In spite of the group delay error being only 1.57 ms, clearly audible ringing is heard when listened in the test. The case has received an average grade of 2.8, which implies a clear degradation in audio quality.

5.2.2 L–R Crossover Response

Despite the group delay error of 2.4 ms, which is clearly audible according to the average grade of 2.9 from the listening test, the time-domain auditory response of the L–R crossover filter shows only a slight difference between original and crossover-processed samples, as Fig. 17 illustrates. This is an example of a time-shift type phase error, different from the FIR crossover ringing, which causes spreading in time.

Studying the off-axis case, the auditory response to the 10-Hz square wave signal with an L–R 32th-order crossover off the main axis is shown in Fig. 18. The delay between drivers is now 0.2 ms. The auditory response to the square wave shows a clearer difference between original and crossover-processed signals. The response curve has a bump on the left side on the crossover channel at

Fig. 16. Auditory analysis responses to 10-Hz square wave signal with 2000th-order FIR crossover filter on three adjacent auditory channels, (a) 3 kHz (crossover channel), (b) 2.75 kHz (lower channel), (c) 3.25 kHz (upper channel). Delay between drivers is 0.02 ms. Crossover-processed signal response is clearly spread in time on auditory channel corresponding to crossover frequency, which is heard as audible ringing in sample. Average grade was 2.8.
3 kHz, while the adjacent channels do not seem to have any abnormalities except time shift. It must be remembered that this sample has a magnitude error greater than 2 dB, which is above the JND limit.

5.3 Auditory Response to Castanets

Another illustration of the ringing phenomenon is presented in Fig. 19. Table 4 lists the parameters of the samples. Fig. 19(a) shows the auditory responses to the castanet signal with a 700th-order FIR crossover, suffering from weird behavior at the tops of the envelope, but still receiving an average grade of 4.4, which means practically no audible distortion.

On the contrary, Fig. 19(b) illustrates the auditory responses to the castanets with a 2000th-order FIR crossover, and strong spreading is noticed. This means audible ringing in the sample, just as with the 10-Hz square wave signal. The smoothed one-third-octave spectra present only

Fig. 17. Auditory analysis responses to 10-Hz square wave signal with 32th-order L–R crossover filter on three auditory channels without original signal. Delay between drivers is 0 ms. Notice how responses differ in time due to changing group delay of L–R crossover, producing audible phase errors.

Fig. 18. Auditory analysis responses to 10-Hz square wave signal with 32th-order L–R crossover filter off axis on three adjacent auditory channels. (a) 3 kHz (crossover channel). (b) 2.75 kHz (lower channel). (c) 3.25 kHz (upper channel). Delay between drivers is 0.2 ms. On auditory channel corresponding to crossover frequency, crossover-processed signal response shows bump at left. Average grade was 3.1.
0.6- and 0.2-dB magnitude errors and 2.25- and 0.93-ms group delay errors for 700th- and 2000th-order FIR crossovers, respectively. So both the group delay error and the magnitude error decrease as the order increases, but the average grade decreases from 4.4 for the 700th-order FIR crossover to 2.3 for the 2000th-order FIR crossover.

**5.4 Conclusions from Auditory Analysis**

1) L–R crossovers show completely different auditory responses (differences in channel group delays, Fig. 17) than high-order FIR crossovers (spreading in time, Fig. 16). Off the main axis, the auditory responses of L–R crossovers show clearer abnormalities (Fig. 18).

2) The ringing caused by the Gibbs phenomenon in high-order FIR crossovers is visible in the auditory responses to 10-Hz square wave and castanet signals, which helps to predict the degradations, while the group delay error values do not imply any audible degradation. The tom-tom drum has so complex waveforms that auditory analysis could not be interpreted easily.

3) Quantitative measures of audio quality are challenging to find because of parallel time- and frequency-domain analyses in hearing. From points 1) and 2) it can be concluded that either group delay error values, magnitude error values, or temporal auditory analysis by inspection reveals the degradation in a signal. The L–R crossovers seem to correlate better with the group delay error values, while high-order FIR crossovers may show qualitatively clearer degradations in the auditory channel envelope responses.

**6 SUMMARY AND CONCLUSIONS**

The experiments and analyses described show that the perception of sound degradations due to crossover filters is complicated. While magnitude and group delay errors exhibit fairly good correlations to perceived degradations in systems with classical crossovers, high-order FIR filters behave differently and require an advanced approach based on auditory (perceptual) modeling.

In this study we have demonstrated that simple auditory modeling can qualitatively reveal how the ringing phenomenon causes degradation in very high-order crossover filters. It remains to be investigated whether a more advanced auditory model can predict subjective test results also quantitatively, possibly combining the magnitude, group delay, and spreading type of degradations in a single model. This remains a challenge to future research.

---

Fig. 19. Auditory analysis responses to castanet signal with (a) 700th- and (b) 2000th-order FIR crossover filters. Delay between drivers is 0.5 ms. On auditory channel corresponding to the crossover frequency, the crossover-processed signal response is flattened from the top, but not much spread. Average grades were 4.4 for order 700 and 2.3 for order 2000.
For digital crossover design we can draw the conclusion that L–R filter orders up to about 8 should be safe in most cases, whereas for FIR filters it is possible to go up to about 600 without problems. Of course the safety limits depend on many factors. For example, in the case of coaxial loudspeakers the delay between driver elements does not vary much with the listening angle, so higher filter orders are possible. In general, going to the safety limit of a filter order may not be well recommended.

7 ACKNOWLEDGMENT

This work was supported in part by the Academy of Finland under project 106956.

8 REFERENCES

Henri Korhola was born in Lappeenranta, Finland, in 1982. He graduated from Helsinki University of Technology, Espoo, Finland, in acoustics and audio signal processing in 2008.

During his studies he worked in a room acoustics consulting company and at the loudspeaker manufacturer Genelec Oy. At present he is with an IT company.

Mr. Korhola’s interests include room acoustics, audio signal processing, and electroacoustics, especially sound reproduction in home theaters. He is also an eager reader of all kinds of books, likes to play snooker and tennis, and searches for improvement in every area of life.

Matti Karjalainen was born in Hankasalmi, Finland, in 1946. He received M.Sc. and Dr.Sc. (Tech.) degrees in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1970 and 1978, respectively.

Since 1980 he has been a professor in acoustics and audio signal processing at the Helsinki University of Technology, Espoo.

Dr. Karjalainen’s main interest is in audio signal processing, such as DSP for sound reproduction and auralization, music DSP, and sound synthesis, as well as perceptually based signal processing. In addition his research activities cover speech processing, perceptual auditory modeling, spatial hearing, DSP hardware, software, and programming environments, as well as various branches of acoustics, including musical acoustics and modeling of musical instruments. He has written 370 scientific and engineering papers and contributed to organizing several conferences and workshops, such as serving as the papers chairman of the AES 16th Conference and as technical chairman of the AES 22nd Conference. He is a fellow and silver medalist of the Audio Engineering Society, and a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, the European Acoustics Association, the International Speech Communication Association, and several Finnish scientific and engineering societies.