The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals*

CHIN-TUAN TAN, AND BRIAN C. J. MOORE, *AES Member*

Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK

AND

NICK ZACHAROV, *AES Member*

Nokia Research Center, Audio-Visual Systems Laboratory, Tampere, Finland

The effect of various types of nonlinear distortion on the perceived quality of speech and music signals was examined. In experiments 1 and 2, "artificial" distortions were used, including hard and soft symmetrical and asymmetrical peak clipping of various amounts, center clipping, and full-range waveform distortion produced by raising the instantaneous absolute value of the waveform to a power ($\neq 1$) while preserving the sign. Subjects were asked to rate the perceived amount of distortion on a ten-point scale (where 1 was most distorted and 10 least distorted). In experiment 1 the distortions were applied to the broadband signals. In experiment 2 the distortions were applied to subbands of the signal. Results were highly consistent across subjects and test sessions. Center clipping and soft clipping had only small effects on the ratings, whereas hard clipping and the full-range distortions had large effects. The subjective ratings were compared to physical measures of distortion based on multitone test signals. A distortion measure, DS, derived from the output spectrum of each nonlinear system in response to a 10-component multitone signal gave high negative correlations with the subjective ratings (correlations were negative as large values of DS were associated with low ratings). A further experiment was conducted using stimuli for which nonlinear distortion was introduced by recording the outputs of real transducers. The output signals were digitally filtered to reduce irregularities in the amplitude–frequency response as far as possible. The results showed moderately strong negative correlations between the subjective ratings and the objective measure DS. It was concluded, that an objective measure of nonlinear distortion based on the use of a multitone signal can predict the perceptual effects of nonlinear distortion reasonably well.

0 INTRODUCTION

All transducers (such as loudspeakers and microphones) and transmission channels (including amplifiers) introduce a certain amount of distortion. The received or reproduced signal is not identical to the original. Distortion can be broadly categorized into two types:

1) *Linear distortion*. This involves changes in the relative amplitudes and phases of the frequency components present in the complex signal. Such changes are typically perceived as changes in timbre or tone quality (coloration) [1]–[6]. In principle, distortion of this type can be compensated (within limits) by linear filtering, and such filtering can be applied either before or after the transducer or transmission channel whose properties are being studied.

2) *Nonlinear distortion*. This involves the introduction of frequency components that were not present in the input signal. The effects of nonlinear distortion are difficult or impossible to compensate by subsequent processing. Signals subjected to nonlinear distortion are often perceived as distorted. It is perhaps unfortunate that the same word is used to describe both the physical process and the subjective impression. The effects of nonlinear distortion may be described as harshness or roughness or in terms of the perception of sounds that were not present in the original signal such as crackles or clicks [7]–[9].

We have been conducting a series of studies with the goal of characterizing, and eventually predicting, the per-

*Manuscript received 2003 July 11; revised 2003 September 2.

petual effects of both linear and nonlinear distortion. We feel that an essential first step is to characterize the perceptual effect of each type of distortion separately. The perceptual effects of linear distortion have been characterized in another paper [10]. The present paper is concerned with the effects of nonlinear distortion on the perceived quality of speech and music. In what follows, the word "distortion" is taken to mean nonlinear distortion and not linear distortion.

Distortion is typically measured using one or two sinusoids as a test signal. If a signal sinusoid is used, the distortion is often specified as total harmonic distortion in percent. When two sinusoids are used, then intermodulation distortion is measured. Again, it is usually expressed as a percentage. These two measures are still used widely for characterizing the performance of loudspeakers, amplifiers, and hearing aids [11]. However, the perception of distortion is not closely related to the percentage of harmonic or intermodulation distortion. Perception depends, among other things, on the frequencies of the distortion products relative to the frequencies of the test signal. For example, if the frequencies of distortion products fall very close to the frequencies of the primary components, the distortion products may be masked by the primary components. In contrast, if the distortion product frequencies are remote from those of the primaries, the distortion products may be heard more easily. In addition, distortion products may interact with primary components to produce temporal effects (amplitude or frequency modulation), and the audibility of these effects depends on the frequency separation and relative phases of the distortion products and primary components. The simple percentages measure of harmonic or intermodulation distortion takes no account of the frequencies and phases of the distortion components relative to the primary components, and hence cannot capture these effects. Furthermore, measures of distortion obtained using sinusoidal test signals may not be appropriate for representing the audibility of distortion in more realistic signals, such as speech and music, whose amplitude fluctuates from moment to moment.

This point has been strongly made in a recent review article [8], [9]. The authors concluded that, "all attempts to find simple quantitative ratios between conventionally measured nonlinear parameters and subjectively detected distortion have not been entirely successful" [8, p. 1027]. They proposed, following the work of Risch [12], that rather than using single-tone or two-tone test signals, distortion should be characterized using a multitone test signal with logarithmically spaced components. They suggested that "the application of the multitone stimulus with its ability to excite numerous distortion products of various orders" might help to provide a link between objective measures of distortion and the subjective perception of distortion. In this paper we explore this link and attempt to predict subjective ratings of distortion using objective measures obtained with a multitone signal.

The papers by Czerwinski et al. [8], [9] provide an extensive historical review of previous studies of the measurement of physical distortion and the perception of distortion, particularly in the evaluation of amplifiers and loudspeakers, and we will not attempt to provide a similar review. However, it is noteworthy that parallel efforts were made in the evaluation of distortion in hearing aids. Kates [13] proposed the use of a comb-filtered noise as a test signal. The proposed noise resembles the multitone signal, except that the peaks are uniformly spaced on a linear frequency scale rather than a logarithmic scale. Several researchers have proposed the use of coherence as a perceptually relevant tool for characterizing distortion [14]–[17]. The coherence is the normalized cross-spectral density between the input and the output, and it has typically been measured using random noise, pseudorandom noise or a maximum-length sequence as the test signal. The measurement of coherence with noise-like signals has some of the advantages of using the multitone signal "because it measures all forms of distortion and not just the harmonic distortion traditionally measured" [18]. However, this approach works well only with time-invariant distortion, and does not give reliable estimates of distortion for hearing aids incorporating automatic gain control. Kates [18] proposed the use of the phase variance of the system transfer function as a perceptually relevant measure of distortion. The phase variance can be estimated from the normalized input–output cross correlation, and it is relatively unaffected by changes in system gain produced by automatic gain control.

Finally it should be noted that methods for assessing the perceived quality of audio signals have been developed for the purpose of evaluating bit-trade reduction systems used in digital coding, complex audio processing, or transmission chains. Examples of such methods include the PEAQ standard ITU-R BS.1387-1 [19] and the PESQ standard ITU-T P.862 [20]. These methods are intended for use only with electrical signals, and not with signals that have been passed through electroacoustic transducers. Their ability to predict degradations in quality produced by nonlinear distortion in transducers has not been assessed. Also, PESQ is intended only for narrow-band speech measurements (300–3400 Hz).

In this paper we first describe two experiments that examine the effects of a variety of "artificial" distortions on the perceived quality of speech and music signals. Some of the distortions, such as hard and soft peak clipping, were meant to resemble distortions that occur in electroacoustic systems such as amplifiers and loudspeakers. Other distortions were not meant to mimic any physical system, but were used to provide a greater variety of types of distortion. The two experiments were based only on static time-invariant nonlinearities. The perceptual data were intended to provide a strong test of methods and models for predicting the perceptibility of distortion. We compare the perceptual data from experiments 1 and 2 with physical measures based on the output of the nonlinear systems in response to multitone signals. A physical measure, DS, based on the spectrum of the output, is developed for predicting perceived distortion. Finally, we present a verification experiment, in which perceptual judgements were obtained of speech and music signals that were subjected to nonlinear distortion produced by various types of transducers.

1 EXPERIMENT 1: WIDE-BAND DISTORTION

In the first experiment the distortions were applied to the entire waveform of the stimulus, resulting in wide-band distortion.

1.1 Types of Distortion

The following types of distortion were applied:

1) Hard symmetrical clipping, with the clipping level set so that the input signal was clipped 0.5, 1, 2, 5, or 10% of the time. The clipping levels were chosen based on distributions of instantaneous amplitude for the entire stimulus, and were determined separately for the speech and music signals (details of these are given later).

2) Hard asymmetrical clipping, with positive peaks only clipped. The clipping level was set so that the input signal was clipped 0.5, 1, 2, 5, or 10% of the time.

3) Soft symmetrical clipping, where the slope of the input–output function decreased for large signal values, but did not become zero. The clipping threshold was defined as the input level at which the output level (for a sinusoid) was 2 dB below the level that would have occurred if the system had remained linear. The clipping was set so that the input signal exceeded the clipping threshold 0.5, 1, 2, 5, or 10% of the time.

4) Soft asymmetrical clipping, with positive peaks only clipped. The clipping was set so that the input signal exceeded the clipping threshold 0.5, 1, 2, 5, or 10% of the time.

5) Center clipping, for which voltages within a certain range around 0 V were set to 0 V. The clipping range was set to 0.5, 1, 2, 5, or 10% of the rms value of the input signal.

6) Full-range distortion produced by raising the instantaneous absolute magnitude of the signal to a power α (not equal to 1) while preserving the sign of the magnitude. Values of α were 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, and 2.

This gave a total of 32 conditions involving a wide range of types and amounts of distortion.

1.2 Test Signals

Two sets of test signals were used, speech and music. Digital representations of the input signals were obtained directly from a CD, using the standard sampling rate of 44 100 Hz. The speech was a concatenation of two sentences, one from a male and one from a female talker, taken from tracks 49 and 50 of the CD "Sound Quality Assessment Material" (SQAM) produced by the European Broadcasting Union.[1] The overall duration of the two sentences, including the brief pause between them, was 3.1 s. The music was a fragment of jazz (piano, bass, and drums) with a relatively constant overall level, taken from a commercial CD (digital recording). The same fragment was used throughout. Its duration was 7.3 s. The speech and music were processed off line, and the processed stimuli were stored on computer disk.

The stimuli were replayed to the listener using a 24-bit Lynx 1 sound card, mounted in a PC. The output of the sound card drove Sennheiser HD580 earphones. The same signal was fed to each earpiece. These earphones have a diffuse-field response, that is, they produce at the eardrum of the listener a similar frequency response as would be obtained listening in a diffuse sound field. Thus their response at the eardrum shows an increase in the frequency range around 3000 Hz, which reflects the resonance normally produced by the concha and meatus. The earphones were calibrated using a KEMAR manikin [21], averaging the results for the "large" and "small" ears. The output of the ear simulator was connected to a Hewlett-Packard 35670A dynamic signal analyzer. The frequency response was compared to the mean of the diffuse-field responses of the human ear measured by Shaw [22], Kuhn [23], and Killion et al. [24]. The response of the earphone was found to match the mean diffuse-field response within ±3.5 dB from 30 to 6000 Hz. Above 6000 Hz the response showed some irregularities, which varied depending on which ear was used in KEMAR and also on the exact positioning of the earphones on the manikin. Additional measurements using a probe microphone (Etymõtic Research ER7C) close to the eardrum of several human individuals showed that the response above 6000 Hz varied from one individual to another, but that the response averages across individuals was close to the diffuse-field response. Such individual variations also occur in the diffuse-field responses of human ears [22]–[24].

We also measured the harmonic and intermodulation distortion produced by the earphone, again using the ear simulator and a Hewlett-Packard 35670A dynamic signal analyzer. For sinusoidal inputs with frequencies from 100 to 6000 Hz, and for sound levels at the output up to 90 dB SPL, the level of the distortion component corresponding to the second harmonic was always at least 70 dB below the level of the primary component, while the level of the distortion component corresponding to the third harmonic was always at least 84 dB down. Other distortion components were not measurable. For various pairs of primary frequencies, f_1 and f_2, in the range 100 to 7000 Hz, and for levels up to 90 dB SPL, the level of the distortion component at $f_2 - f_1$ was always unmeasurable. For the same conditions, the level of the distortion component at $2f_1 - f_2$ was always at least 73 dB lower in level than the primary tones, and was usually unmeasurable. No other distortion components were measurable. We conclude that the distortion produced by the earphones was probably below the audible limit [7]–[9], [25], [26]. The distortion was certainly well below that introduced deliberately into our stimuli.

The overall level of each distorted signal was adjusted digitally prior to digital–analog conversion so as to give roughly a constant loudness of 86.4 phons (binaural listening). The required adjustment was calculated using the loudness model of Moore et al. [27]. The calculations took into account the diffuse-field response of the earphones.

1.3 Experimental Method

In a given test sequence of 32 stimuli, a listener was tested using either speech or music. Stimuli subjected to the different types of nonlinear distortion were presented in a randomized order. After each stimulus presentation there was a pause, during which the listener was required

[1] www.ebu.ch; materials also available from http://sound.media.mit.edu/mpeg4/audio/sqam/.

to rate the perceived quality on a 10-point scale where 10 indicates "clean, completely undistorted" and 1 represents "very distorted" The response categories were displayed on the computer screen, and subjects responded using the mouse to click on their category of choice. The computer waited indefinitely until a response was made. The next stimulus was presented approximately 1 s after a response was made. It should be noted that asking subjects to rate quality in this way is different from asking subjects to judge the difference between the distorted signal and the original. Sometimes specific types of distortion can actually increase sound quality [28]. However, distortion of the type used here introduces intermodulation components that are often not harmonically related to the original signal and which always result in a degradation of sound quality.

To illustrate the meaning of the descriptors for the categories, before the experiment started, samples were presented of undistorted signals; these were described as examples of category 10. Similarly, samples were presented with large amounts of distortion (full-range distortion with $\alpha = 0.5$ or 2); these were described as category 1. Each subject was tested in two sessions on different days. In each session the test was conducted once using speech and once using music. The repeated measurement for each type of stimulus allowed us to assess how consistent the responses of each subject were. A session typically lasted about one hour.

1.4 Subjects

Ten subjects were tested. None had any history of hearing disorders and all had audiometric thresholds better than or equal to 20 dB HL in both ears at all audiometric frequencies from 250 to 8000 Hz. Their ages ranged from 15 to 35 years (mean 24, standard deviation 6). Subjects were paid for their participation.

2 RESULTS

1.4 Consistency across Sessions and Subjects

The results for each subject generally showed a very similar pattern across the two test sessions for a given type of signal (speech or music). The overall consistency across test sessions was assessed by calculating the mean score across subjects for each condition and stimulus type, separately for each session, and then calculating the correlation of the scores for the 32 conditions across sessions. The correlations obtained in this way were 0.988 for the speech stimuli and 0.985 for the music stimuli. The very high correlations indicate a high degree of consistency of the group mean scores across test sessions.

The pattern of results was also very consistent across subjects. To assess the degree of consistency across subjects, we calculated the mean score for each subject and each condition across the two sessions. We also calculated the mean score across subjects for each condition and stimulus type, including the data for both sessions. Then, for each subject in turn, we calculated the correlation between the scores for that individual subject and mean scores, over the 32 conditions. The higher the correlation, the more closely the pattern of scores for a given subject resembles that for the group as a whole. The resulting correlations are shown in Table 1, separately for music and for speech stimuli. The resulting correlations are high, indicating a high degree of consistency across subjects. The standard deviation (SD) of the ratings across subjects for a given condition was typically about 1.3 scale units (standard error, SE, about 0.4 scale unit).

For each type of distortion (such as hard symmetrical clipping) an ANOVA was conducted on the ratings with factors subject and amount of distortion (such as clipping 0.5, 1, 2, 5, or 10% of the time). This was done separately for the speech and music stimuli. The scores for the two test sessions were treated as replications. For each type of distortion, there was a highly significant effect of the amount of distortion ($p < 0.001$); ratings decreased in an orderly way with increasing distortion. For the full-range distortion, ratings decreased as α was made more different from 1 (either larger or smaller than 1). The effect of subject was sometimes significant, indicating that some subjects gave lower or higher overall ratings than others. There were only two cases where the interaction of subject and amount of distortion was significant, indicating that the *pattern* of results differed across subjects. These cases were for speech stimuli with full-range distortion ($F(54, 70) = 1.84$, $p = 0.008$) and music stimuli with asymmetrical hard clipping ($F(36, 50) = 1.88$, $p = 0.019$). In both cases the interaction term accounted for less than 2% of the variance in the data. We conclude that the ratings are mainly determined by the amount of distortion in the stimuli, and that individual differences in the pattern of ratings are small. In what follows, we focus on the mean ratings.

2.2 Mean Ratings

The mean ratings across subjects are shown in Figs. 1 and 2 for the speech and music stimuli, respectively. In general the data are very orderly. For a given type of distortion (such as symmetrical peak clipping), the ratings decrease monotonically with increasing physical distortion.

Even for small amounts of distortion, the mean rating was never above 9.2 for speech and 9.0 for music. This reflects the fact that subjects tend to avoid the extremes of the available range of responses when making sub-

Table 1. Correlation of mean ratings across sessions for each individual subject with mean ratings across subjects, for experiment 1.

Subject	1	2	3	4	5	6	7	8	9	10
Speech	0.94	0.95	0.94	0.97	0.97	0.96	0.96	0.97	0.98	0.91
Music	0.96	0.95	0.94	0.97	0.97	0.98	0.97	0.98	0.96	0.95

jective judgments [29]. The lowest ratings were obtained for the full-range modification with $\alpha = 0.5$ or 2; the mean ratings were close to 1 for these cases. Hard symmetrical clipping 10% of the time also gave low ratings, around 3.5 for speech and 3.1 for music. Soft clipping and center clipping up to 2% had little effect on perceived quality.

3 EXPERIMENT 2: BAND-LIMITED DISTORTION

In experiment 2 the distortion was introduced in a frequency-specific way, so as to gain insight into the relative importance of distortion in different frequency regions.

3.1 Stimuli

In one set of conditions, referred to as prefiltering, the stimuli were filtered into four frequency bands and the distortion was applied to the waveform at the output of one of the filters only. The four bands covered the following frequency ranges: 0–606 Hz, 606–1973 Hz, 1973–5583 Hz, and 5583–22 050 Hz. The outputs of the filters were then recombined. This meant that the distortion originated in a specific frequency band, but the resulting distortion components were allowed to spread to other bands, as might occur in a multiway loudspeaker system in which one transducer introduced distortion. The filters were designed so that the filtering and recombination did

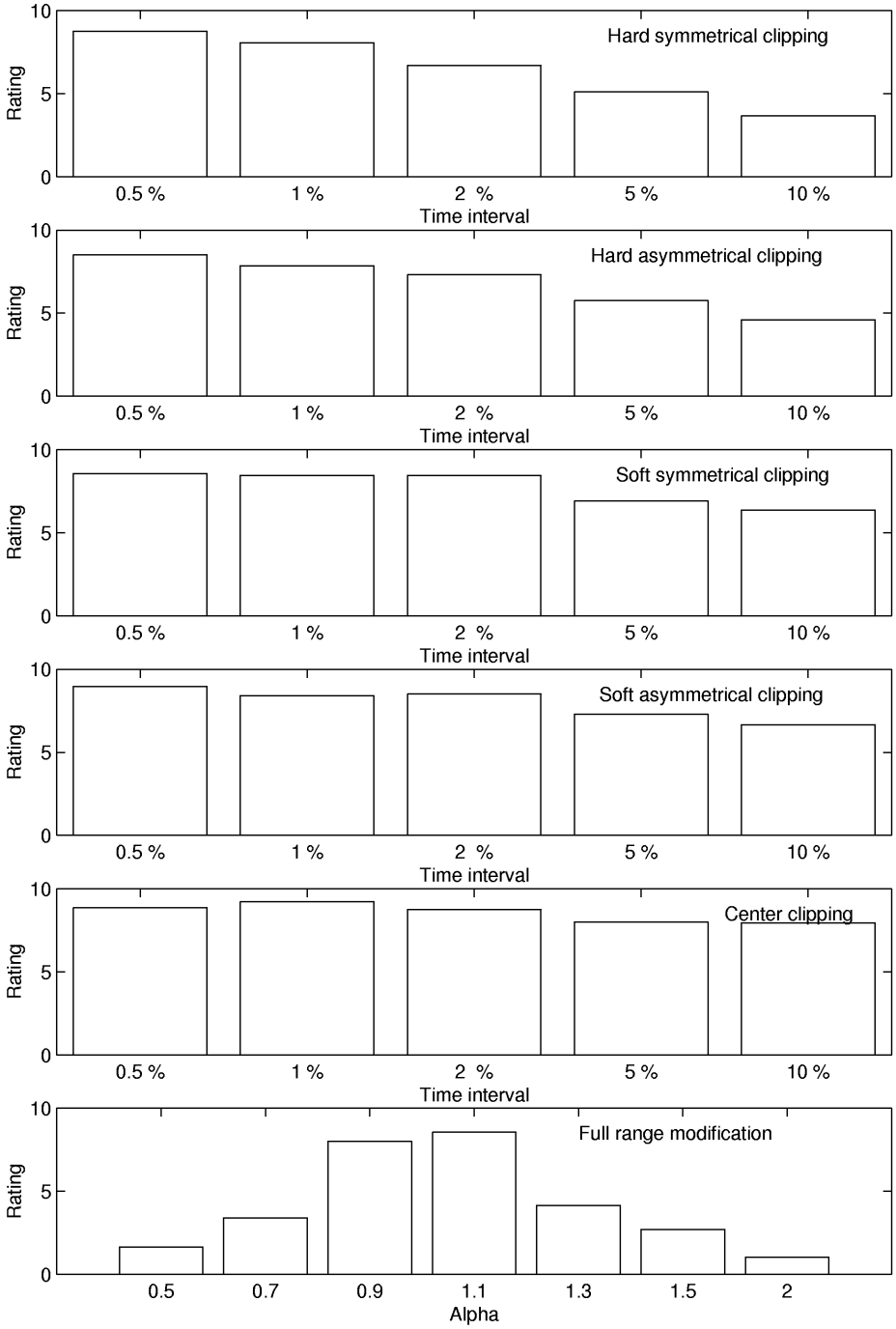

Fig. 1. Results of experiment 1, showing mean ratings for each type of distortion for speech stimulus.

not lead to significant changes in the amplitude–frequency response. The following prefiltering band-specific distortions were used.

1) Hard symmetrical clipping, with the clipping level set so that the input signal was clipped 2 or 10% of the time.

2) Soft symmetrical clipping, with the clipping set so that the input signal exceeded the clipping threshold 5 or 20% of the time.

3) Full-range distortion with $\alpha = 0.7$ or 1.3.

In the second set of conditions, the distortion was applied to the waveform at the output of one of the filters, as before, but then the distorted waveform was postfiltered using the same filter. For example, if the distortion was applied to the output of the filter covering the range 606–1973 Hz, the distorted waveform was subsequently bandpass filtered over the range 606–1973 before the outputs of the different filters were recombined. This restricted the distortion components to a specific frequency region. Such distortion would occur only rarely in a real nonlinear system, but it was used here to allow us to gain some insight into the relative importance of the distortion components in different frequency regions. The pre- and postfiltering band-specific distortions were the same as those described in 1) to 3) for prefiltering only. In addition, we included some distortions applied to the wide-band signal, similar to those used in experiment 1. The wide-band distortions were as follows

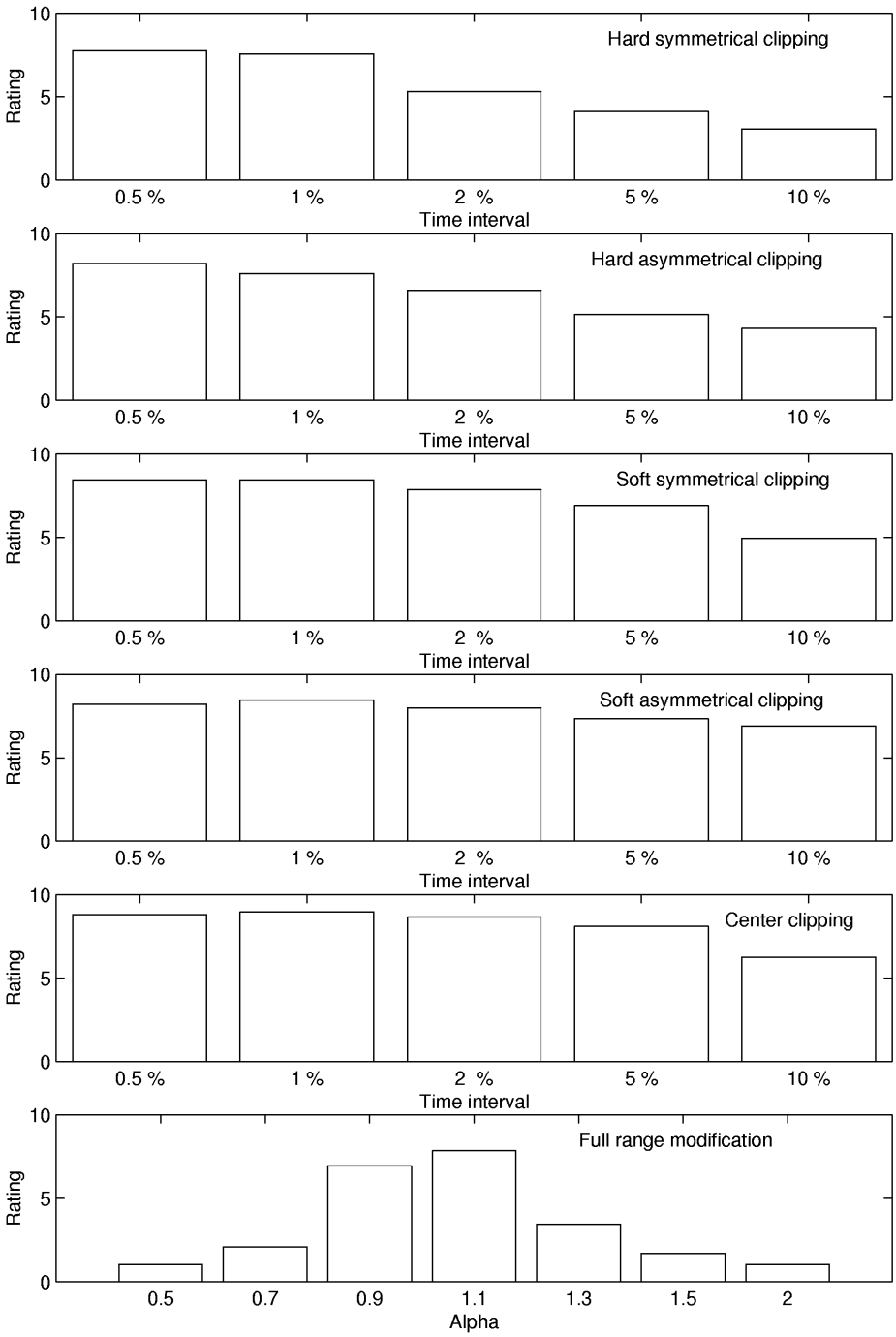

Fig. 2. As Fig. 1, but for music stimulus.

1) Hard symmetrical clipping, with the clipping level set so that the input signal was clipped 2 or 5% of the time.

2) Soft symmetrical clipping, with the clipping set so that the input signal exceeded the clipping threshold 5 or 20% of the time.

3) Full-range distortion with $\alpha = 0.5$, 0.7, 1.3, and 2. We also included a condition with no distortion.

In summary, there were 24 conditions involving prefiltering (six types of distortion applied in each of four passbands), 24 conditions involving pre- and postfiltering (six types of distortion applied in each of four passbands), eight conditions involving distortions applied to the broad-band signal, and one condition with no distortion, giving a total of 57 conditions.

3.2 Subjects and Procedure

Eleven normally hearing subjects were tested. None had any history of hearing disorders and all had audiometric thresholds better than or equal to 20 dB HL in both ears at all audiometric frequencies from 250 to 8000 Hz. Their ages ranged from 20 to 27 years (mean 23, standard deviation 2). Subjects were paid for their participation. The procedure was the same as described for experiment 1.

4 RESULTS

4.1 Consistency across Sessions and Subjects

As before, the consistency across test sessions was assessed by calculating the mean score across subjects for each condition and stimulus type, separately for each session, and then calculating the correlation of the scores for the 57 conditions across sessions. The correlations obtained in this way were 0.976 for the speech stimuli and 0.966 for the music stimuli. The very high correlations indicate a high degree of consistency of the group means across test sessions.

The pattern of results was also consistent across subjects. Correlations between individual scores and mean scores, calculated as for experiment 1, are shown in Table 2. The correlations are high, indicating a high degree of consistency across subjects, although the consistency is not quite as high as for experiment 1. The SD of the ratings across subjects for a given condition was typically about 1.6 scale units (SE about 0.5 scale unit), which is slightly higher than for experiment 1. This may reflect individual variability in the relative importance of different frequency regions.

As before, ANOVAs were conducted on the ratings with subject and amount of distortion as factors. For the distortions involving filtering, the type of filtering was included as a factor (prefiltering alone, or pre- and postfiltering) and the center frequency of the band where the distortion was introduced was also a factor. ANOVAs were calculated separately for the speech and music stimuli. We consider in this section only the outcomes relating to individual differences; other outcomes are discussed in the following section.

For the broad-band distortions the effect of subject was sometimes significant, indicating that some subjects gave lower or higher overall ratings than others. The interaction of subject and amount of distortion was not significant in any case, indicating that the *pattern* of results did not differ across subjects. For the distortions involving filtering, the effect of subject was sometimes significant, and there were a few cases where subject interacted with other factors, indicating a difference in the pattern of results across subjects. These cases were as follows.

1) For speech stimuli with hard symmetrical clipping, there was a significant interaction of subject with type of filtering (pre or pre and post); $F(10, 176) = 2.8$, $p = 0.003$.

2) For speech stimuli with full-range distortion, there was a significant interaction of subject with α, $F(10, 176) = 2.64$, $p = 0.005$, and of subject with band center frequency $F(30, 176) = 2.19$ $p < 0.001$.

3) For music stimuli with hard symmetrical clipping, there was a significant interaction of subject with band center frequency, $F(30, 176) = 1.79$ $p = 0.011$.

4) For music stimuli with full-range distortion, there was a significant interaction of subject with α, $F(10, 176) = 3.31$, $p < 0.001$, and of subject with band center frequency, $F(30, 176) = 2.26$, $p < 0.001$.

These interactions never accounted for more than 1% of the variance in the data, and in three out of four cases they accounted for less than 0.5% of the variance. As before, we conclude that the ratings are mainly determined by the amount and type of distortion in the stimuli, and that individual differences in the patterns of ratings are small.

4.2 Mean Ratings

The mean ratings for the distorted stimuli are shown in Figs. 3–8. The ratings for the broad-band distortions (Figs. 3 and 6) are similar to those obtained in experiment 1. The ANOVAs showed that ratings always decreased significantly with increasing amounts of distortion ($p < 0.001$ in all cases). For the full-range distortion, ratings decreased as α was made more different from 1 (becoming either larger or smaller than 1). The general pattern of the results was similar for the speech stimuli and for the music stimuli. For the prefiltering distortions (Figs. 4 and 7) the ratings were generally lower when the distortion was introduced into the lower frequency bands (but the distortion was allowed to spread to other frequencies) than

Table 2. Correlation of mean ratings across sessions for each individual subject with mean ratings across subjects, for experiment 2.

Subject	1	2	3	4	5	6	7	8	9	10	11
Speech	0.92	0.91	0.92	0.94	0.92	0.92	0.94	0.96	0.93	0.96	0.91
Music	0.89	0.91	0.93	0.93	0.94	0.92	0.96	0.95	0.89	0.93	0.89

when the distortion was introduced into the higher frequency bands. However, for the pre- and postfiltering distortions (Figs. 5 and 8) the effects were much more uniform across the different frequency bands, suggesting that the distortion components within each band were roughly equally important. These patterns are revealed in the ANOVAs as significant overall effects of band center frequency ($p < 0.001$ for both speech and music for all types of distortion) and significant interactions between type of filtering (pre or pre and post) and band center frequency; the significance level was $p < 0.001$ for all types of distortion for both speech and music. An exception to the trend of roughly equal importance of each frequency band for pre- and postfiltering was for the full-range modification with $\alpha = 0.7$. For this case the ratings for pre- and postfiltering were lower for the two middle bands than for the lowest or highest bands. Post hoc tests, based on Fisher's least significant differences test, showed that the mean ratings for the two middle bands were significantly lower than the mean ratings for the lowest and highest bands ($p < 0.001$ in all cases).

The fact that, for prefiltering distortions, the ratings were generally lower when the distortion was introduced into the lower frequency bands can probably be attributed to the physical effects of the distortions. The distortions introduced into the lower bands would have led to harmonic distortion products that covered a wide frequency range, including the upper bands. When pre- and postfiltering were used, the out-of-band distortion products were removed, and this led to higher overall ratings than for prefiltering alone, and also to more uniform ratings across frequency bands.

5 PHYSICAL MEASURES OF DISTORTION USING MULTITONE SIGNALS

The multitone signals used here had properties similar to those described by Czerwinski et al. [8], [9], including the logarithmic spacing of components, as recommended by them. However, their multitone signals had a lowest component of 1000 Hz and a highest component of 10 000 Hz. Such signals would not have revealed distortions in some of the conditions of experiments 2, for which the distortion originated in the lowest frequency band. Therefore we decided to use multitone signals whose components spanned the range from 50 to 15 000 Hz. The root-mean-square (rms) value of each multitone input signal was set equal to the input level of the speech and music signals used in experiments 1 and 2. We used a variety of multitone signals, differing in their number of components, which ranged from 5 to 60. In all cases the components were uniformly spaced on a logarithmic scale over the range 50 to 15 000 Hz. For example, for a multitone complex with 20 components the frequency ratio between

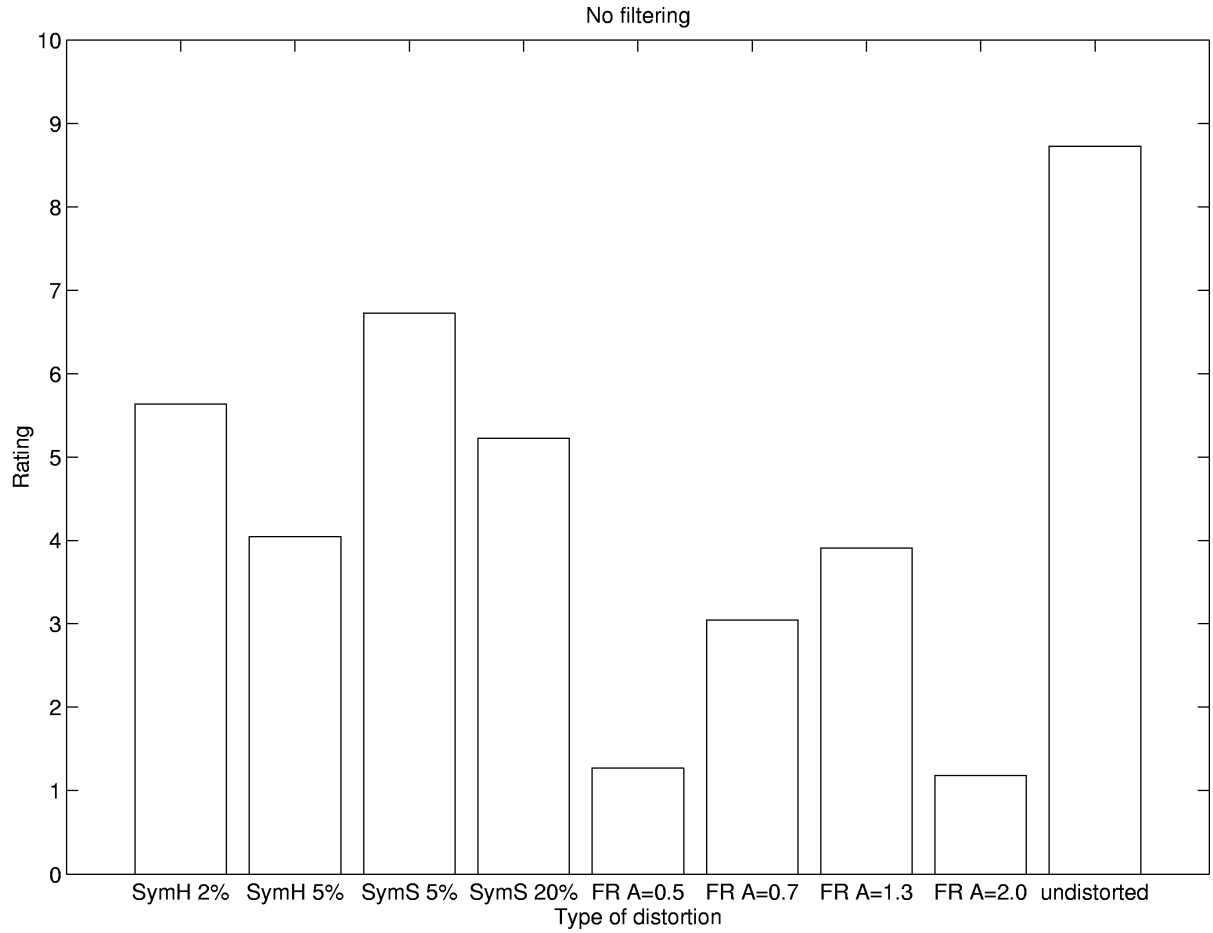

Fig. 3. Mean results of experiment 2 for broad-band distortions for speech stimulus. SymH—symmetrical hard clipping; SymS—symmetrical soft clipping; FR—full-range waveform distortion. Extreme right-hand column shows mean rating for undistorted stimulus.

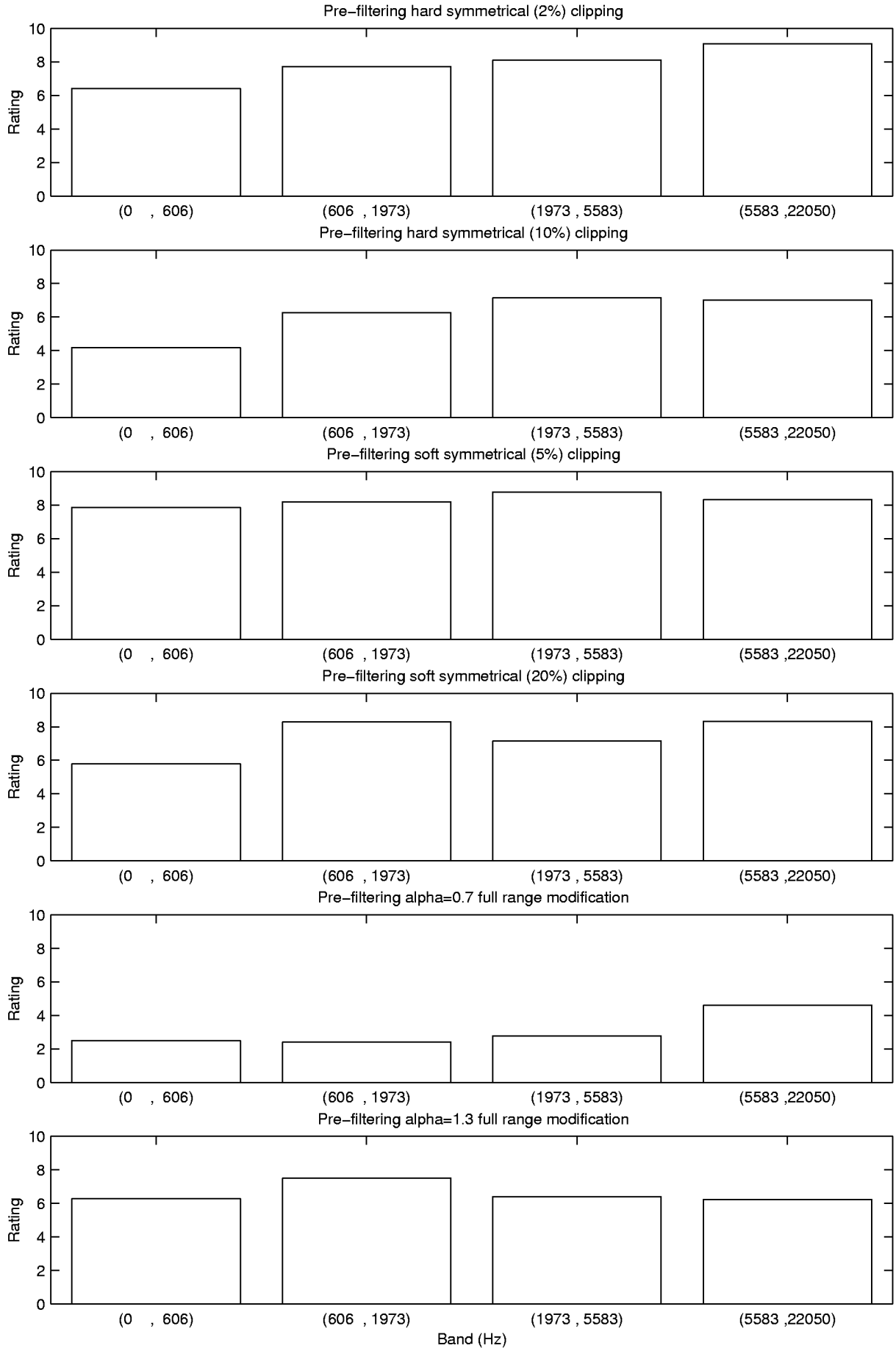

Fig. 4. Mean results of experiment 2 for prefiltering distortions for speech stimulus.

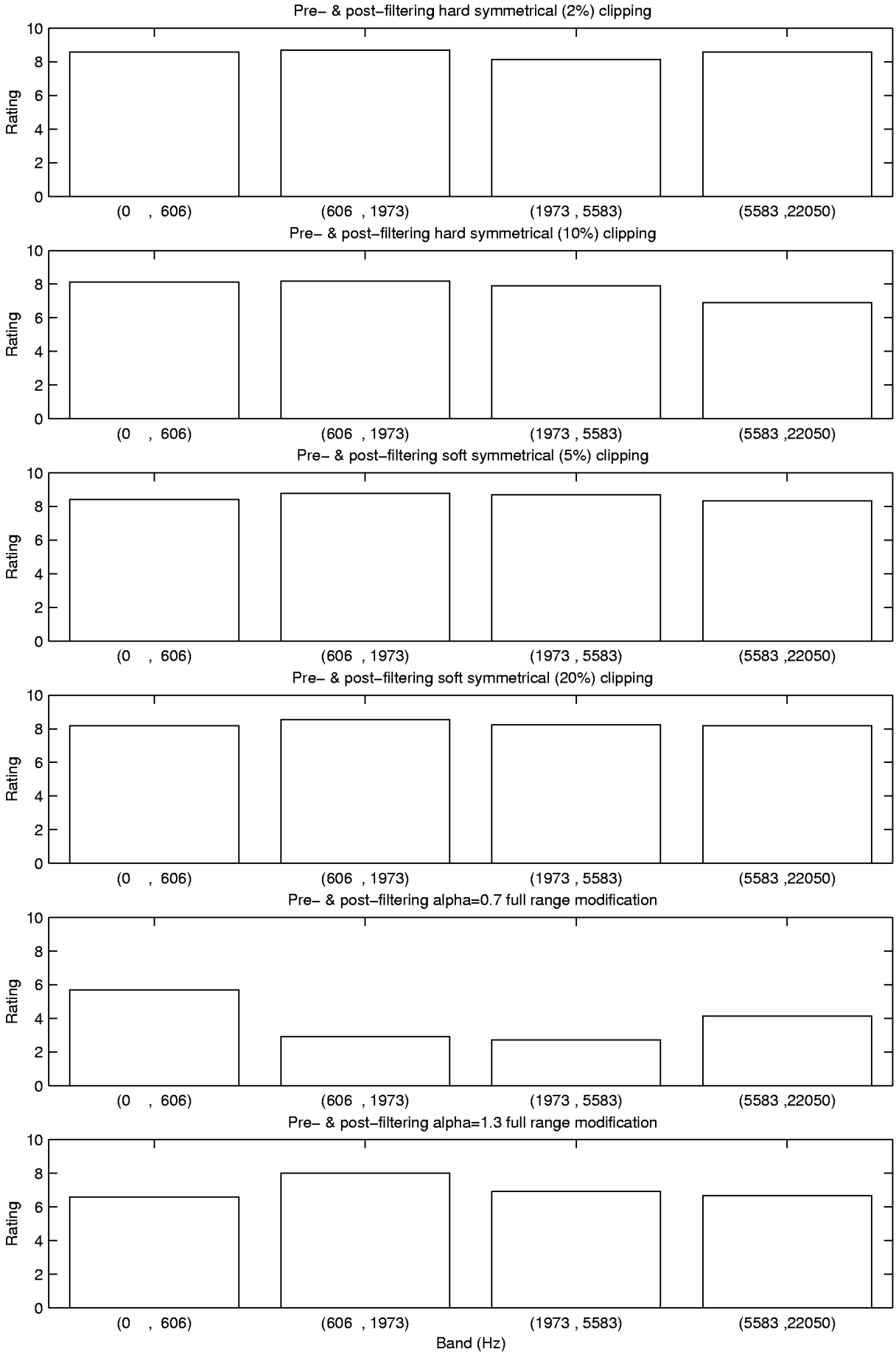

Fig. 5. Mean results of experiment 2 for pre- and postfiltering distortions for speech stimulus.

successive components was 1.35. For initial analyses the duration of each multitone signal was 1 s.

A plot of the long-term spectrum of the output of a nonlinear system in response to a multitone signal gives an immediate visual indication of the extent of distortion [8], [9]. This is illustrated in Fig. 9, which is based on a 10-component multitone signal. Fig. 9 (a) shows the spectrum of the input signal, and Fig. 9 (b) shows the output of the broad-band full-range nonlinear system from experiments 1 and 2, for which the instantaneous amplitude of the signal was raised to the power of 0.5. This system received very low ratings for both speech and music. The severe distortion introduced by this system is readily apparent. Fig. 9 (c) shows the output of the system used in experiment 2 with 2% hard clipping in the highest frequency band, using pre- and postfiltering. This system received high ratings for both speech and music. The distortion apparent in the output spectrum is correspondingly rather low and is restricted to high frequencies.

Although these figures are informative, it is not obvious how to use the output spectrum to give a quantitative estimate of the perceived quality of the distorted signals. One approach would be to calculate excitation patterns evoked in the auditory system by the input and output spectra, and to use the differences between the excitation patterns as an estimate of the perceptual difference [6]. However, this approach is computationally intensive. Also, an approach based on excitation patterns would not work well in cases where the device under test produced linear distortion (that is, the device had a nonflat frequency response) as well as nonlinear distortion. Here we adopted a simpler method of quantifying the differences between the input and output spectra, but a method that is still related to the spectral resolution of the auditory system.

Initially the input and output were time aligned, to compensate for any time delay caused by the nonlinear processing. Both the input and output of each nonlinear system were analyzed in a series of successive nonoverlapping 30-ms frames. The power spectrum of each frame was determined using a 1323-point discrete Fourier transform (DFT). The overall magnitude of the output spectrum was scaled so that the major peaks had the same magnitude as for the input spectrum. The bins in the DFT were grouped into nonoverlapping frequency bands, each of which was 1 ERB_N wide, where ERB_N denotes the mean equivalent rectangular bandwidth of the auditory filter for young normally hearing subjects at moderate sound level [6], [30]. The ERB_N is conceptually similar to the traditional critical bandwidth [31] but differs somewhat in numerical values. The powers of the individual bins within each frequency band were summed to give the overall power in each 1-ERB_N band, and the power was converted to decibels. For each frame the difference in level in each 1-ERB_N band between the input and output was calculated, and the absolute value of the difference was summed across bands. This gives a perceptually

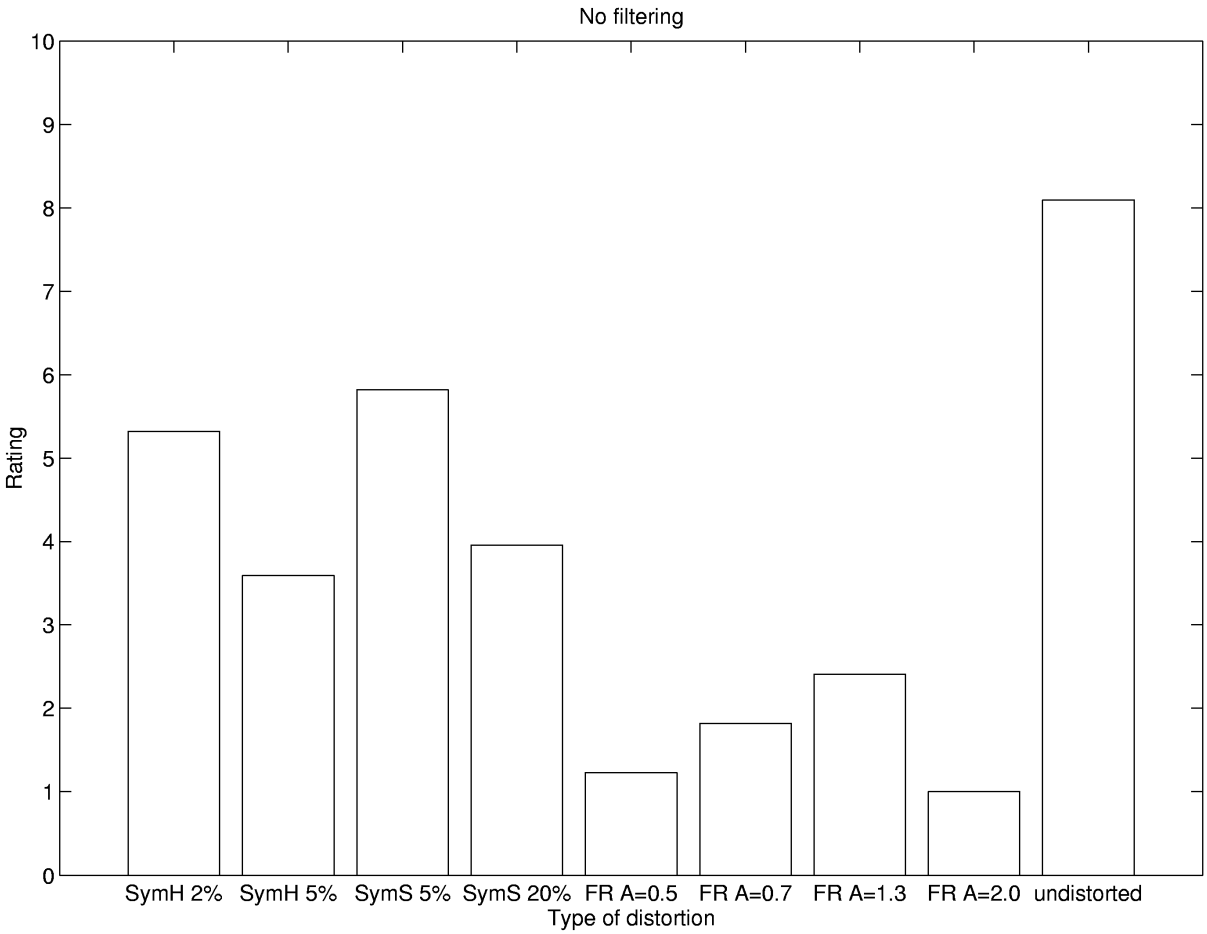

Fig. 6. As Fig. 3, but for music stimulus.

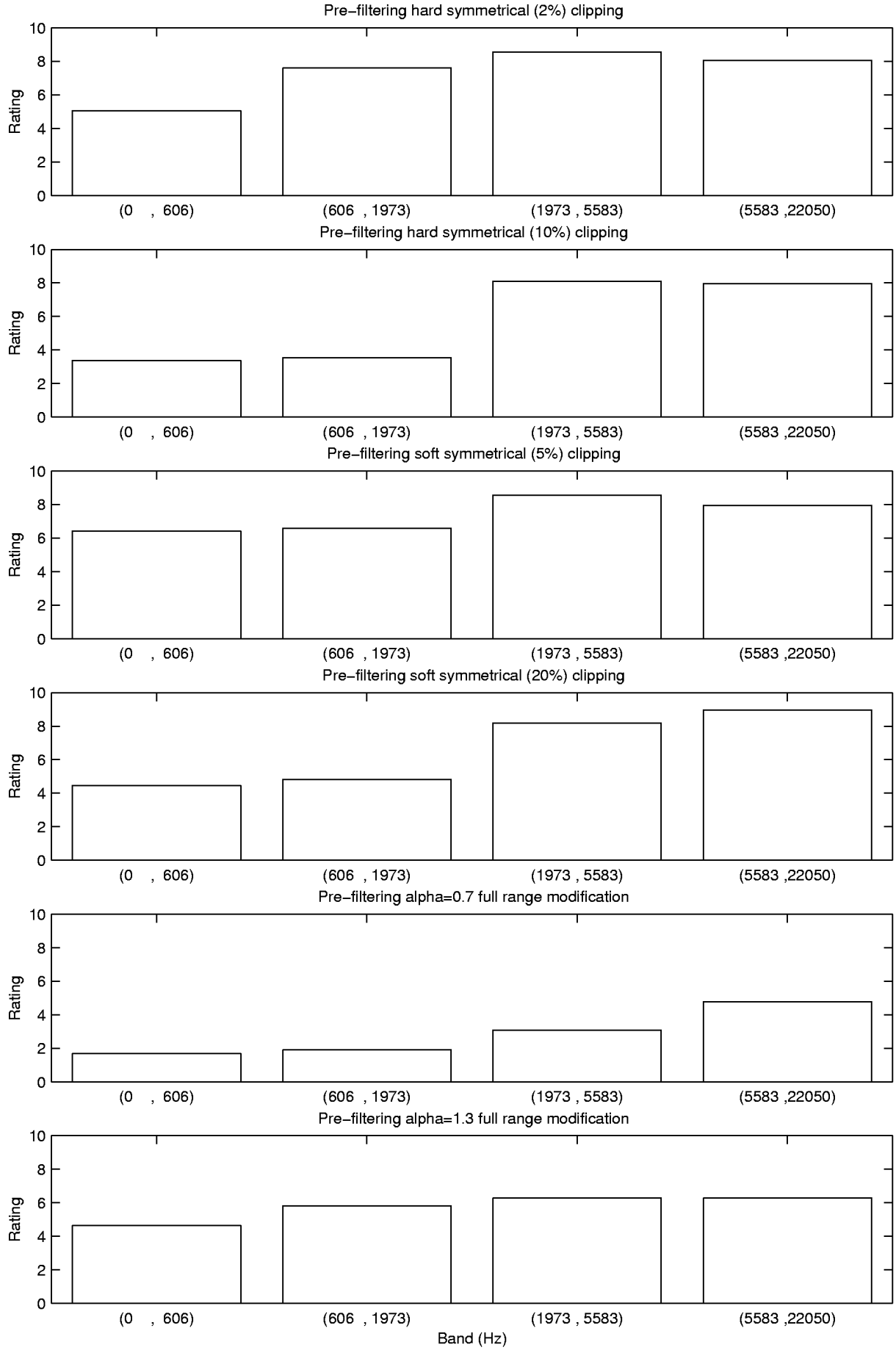

Fig. 7. As Fig. 4, but for music stimulus.

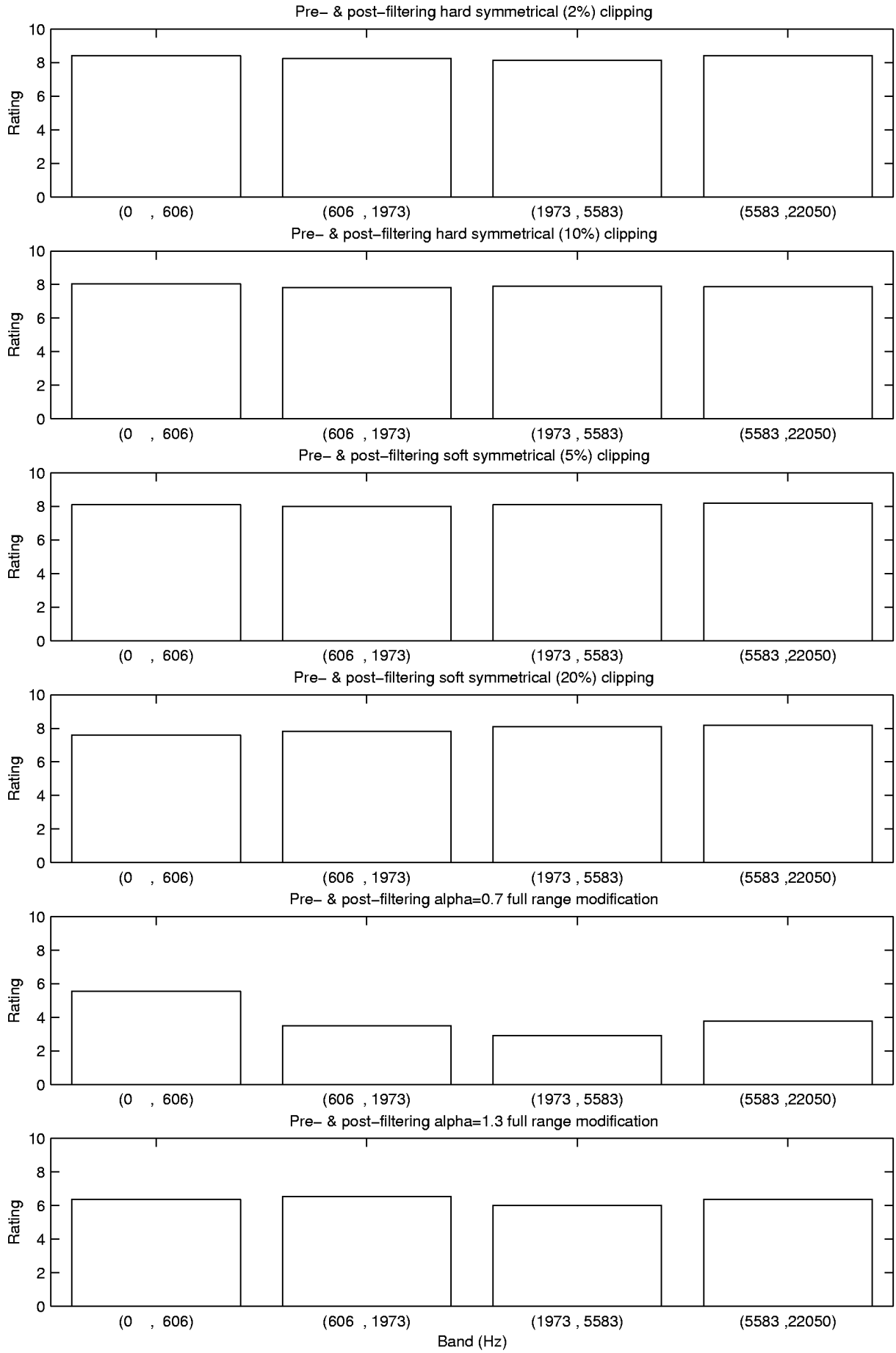

Fig. 8. As Fig. 5, but for music stimulus.

relevant measure of the difference in spectrum between the input and output, which we will call the distortion score, DS. The values of DS were averaged across all 30-ms frames to give an overall measure of distortion.

Several parameters were varied to explore their effect on the relationship between the DS values and the subjective ratings. These included the number of components in the multitone signal, the relative phases of the components in the multitone signal, the duration of the multitone signal, and the effect of windowing the 30-ms segments prior to calculating the DFT.

For each of the multitone signals (with 5–60 components) the DS values were calculated for each nonlinear system used in experiments 1 and 2, and the correlation was calculated between the DS values and the subjective ratings. This was done separately for the ratings of the speech and music signals. As expected, the correlations were always negative, as a large value of DS implies a large amount of distortion and therefore a low rating. For both experiment 1 and experiment 2, the correlations

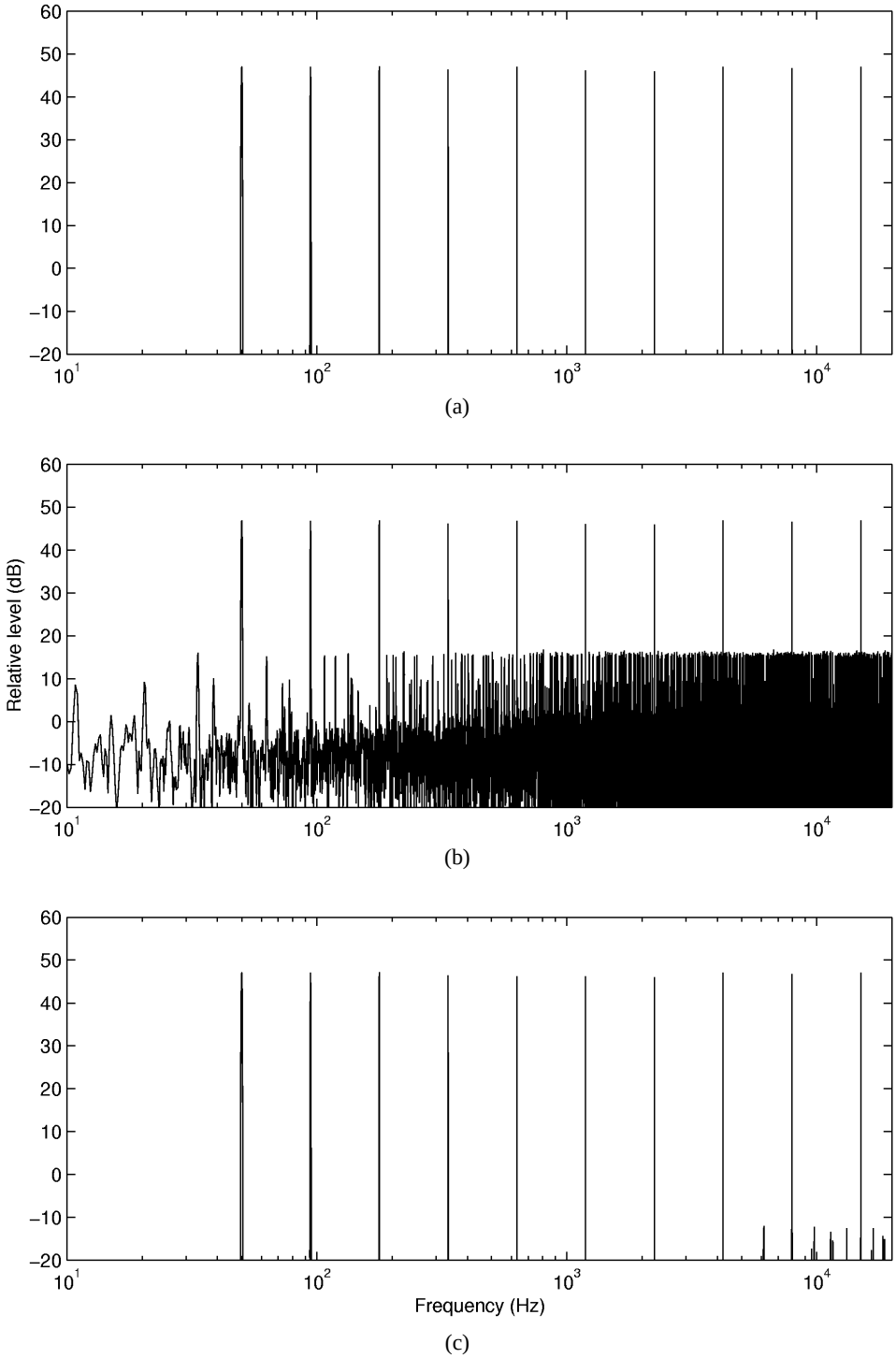

Fig. 9. (a) Spectrum of 10-component multitone signal. (b) Output spectrum of nonlinear system producing severe broad-band distortion. (c) Output spectrum of nonlinear system producing mild distortion at high frequencies only.

obtained in this way were maximal for a multitone signal containing 10 components, and decreased slightly in absolute value when the number of components was decreased to 5 or was 20 or above. Hence we decided to focus on the results obtained using multitone signals with 10 components, which gave strong negative correlations for both experiments. For further evaluations we used several different forms of the multitone signal, differing in the (randomly selected) starting phases of the components. We found that the starting phase had only a small effect on the outcome, but we chose the set of phases that gave the highest negative correlation between the subjective ratings and the DS values. The values used are shown in Table 3.

Increasing the duration of the multitone signal beyond 1 s did not increase the absolute value of the correlations, so we restricted the duration to 1 s. Windowing the 30-ms segments of the output signal prior to calculating the DFT had the effect of reducing spectral splatter from the major components of the signal (those present at the input), thus making it possible to measure the effects of distortion components at lower levels than when no windowing was used. However, we found that the windowing made the absolute values of the correlations smaller rather than larger. This may have happened because distortion components at very low levels would not be perceptually relevant, as they would be masked by the primary components or would be below absolute threshold. Hence for the analyses presented next, the segments were not windowed prior to calculating the DFT.

Fig.10 shows scatter plots of the relationship between the DS values and the subjective ratings for experiment 1. The (inverse) correlation between the two quantities is high, at -0.98 ($p < 0.001$) for speech and -0.97 ($p < 0.001$) for music. Fig. 11 shows similar scatter plots for experiment 2. Here the scatter is greater, but the correlations are still reasonably high, namely -0.87 ($p < 0.001$) for speech and -0.80 ($p < 0.001$) for music. We conclude that the physical measure DS is closely related to perceived distortion, at least for the artificial types of distortion used in experiments 1 and 2.

Table 3. Frequencies and starting phases of 10-component multitone signal that gave the highest (negative) correlation between distortion measure DS and ratings obtained.

Hz	rad
50	2.549
94	5.878
178	5.761
335	2.578
631	5.615
1189	0.364
2241	2.217
4223	5.109
7959	0.062
15000	0.873

6 VERIFICATION EXPERIMENT

6.1 Stimuli

To assess whether the DS measure gave reasonable predictions of the perceived quality of signals subjected to the distortion produced by real transducers, we conducted a third perceptual experiment. The same speech and music signals as before were used as input to the nonlinear systems under test. A few of the nonlinear systems used earlier were included, namely:

1) Hard symmetrical clipping of the broad-band signal, with the clipping level set so that the input signal was clipped 2 or 5 % of the time.

2) Soft symmetrical clipping of the broad-band signal, with the clipping set so that the input signal exceeded the clipping threshold 5 or 20 % of the time.

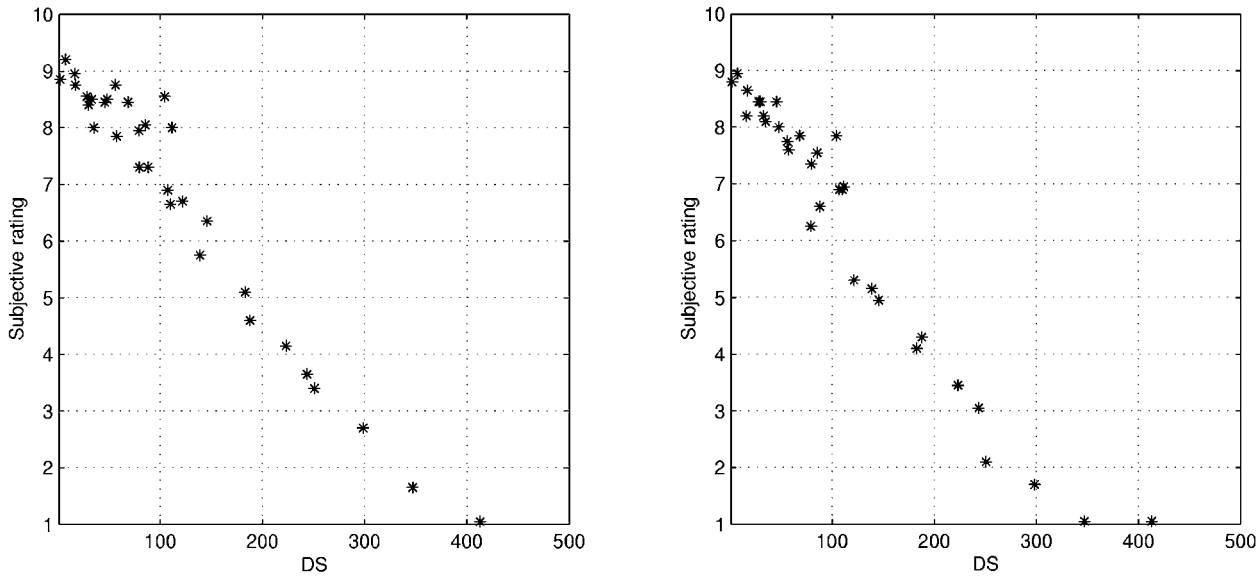

Fig. 10. Scatter plots of mean subjective ratings form experiment 1, plotted against distortion measure DS. (a) Results for speech stimulus. (b) Results for music stimulus. Each point represents a specific form of broad-band distortion.

3) Full-range distortion of the broad-band signal with $\alpha = 0.5$, 0.7, 1.3, and 2.

4) Prefiltering band-specific distortion with $\alpha = 0.7$ and 1.3 in the bands 606–1973 Hz and 1973–5583 Hz.

The other nonlinear systems involved four different compact electrodynamic transducers (13–20 mm in diameter) mounted in a variety of acoustic structures, selected to be representative of earpieces and so-called integrated hands-free designs often encountered in telecommunications terminal devices or accessories. Both commercial products and acoustic mockups were used. These included closed-box and ported acoustics, with a range of back volumes (1–5 cm^3), a number of different front volumes, and different sound port arrangements. The total number of systems used was 15.

The test signals (sampling rate 41 000 Hz) were fed to the test systems from a PC via a Lexicon CD-2, acting as a DAC, and via a Lab Gruppen LAB300 amplifier. Each transducer was driven at two power levels, one of which was below the official rated power handling of the transducer and the other of which was 7–20 dB above the rated power. This was done to ensure the generation of both weak and strong nonlinearities, without causing permanent damage to the transducers. The output of each transducer was measured in the free field using a G.R.A.S. 40AF 0.5-in free-field microphone, a G.R.A.S. 2AA preamplifier, and a Brüel & Kjær Nexus conditioning amplifier. The microphone was located 250 mm from the sound outlet of the device or on the axis of the loudspeaker, depending on the test configuration. The measured signal was recorded to the PC hard disk (sampling rate 44 100 Hz) via a Tascam DA30 (acting as ADC) and an RME Digital 96/8 sound card.

In summary, there were eight conditions involving artificial distortion of the broad-band signal, four conditions involving artificial distortion with prefiltering (two bands times two values of α), and 30 conditions involving real transducers (15 systems, each driven at two levels). We also included a condition with no distortion. The total number of conditions was 43.

One problem in using real transducers is that they introduce linear distortion (frequency response irregularities) as well as nonlinear distortion. Our interest in the present study was to characterize the perceptual effects of the nonlinear distortion without any confounding effect of linear distortion. Therefore the recorded outputs of the real transducers were each digitally filtered so that the long-term spectrum of the output matched that of the input as closely as possible. In practice this could not be done over a very wide frequency range, as some of the transducers showed a considerable rolloff in their amplitudes response at low frequencies. Therefore to match the amplitude–frequency response of all systems as closely as possible, all stimuli were bandpass filtered between 301 and 15 900 Hz. This was done both for the stimuli recorded via real transducers and for those generated using artificial distortion.

6.2 Subjects and Procedure

Nine normally hearing subjects were tested. None had any history of hearing disorders and all had audiometric thresholds better than or equal to 20 dB HL in both ears at all audiometric frequencies from 250 to 8000 Hz. Their ages ranged from 20 to 31 years (mean 27, standard deviation 4). Subjects were paid for their participation. The procedure was the same as described for experiment 1.

6.3 Consistency across Sessions and Subjects

As before, the consistency across test sessions was assessed by calculating the mean score across subjects for each condition and stimulus type, separately for each session, and then calculating the correlation of the scores for the 43 conditions across sessions. The correlations obtained in this way were 0.94 for the speech stimuli and 0.91 for the music stimuli. The high correlations indicate

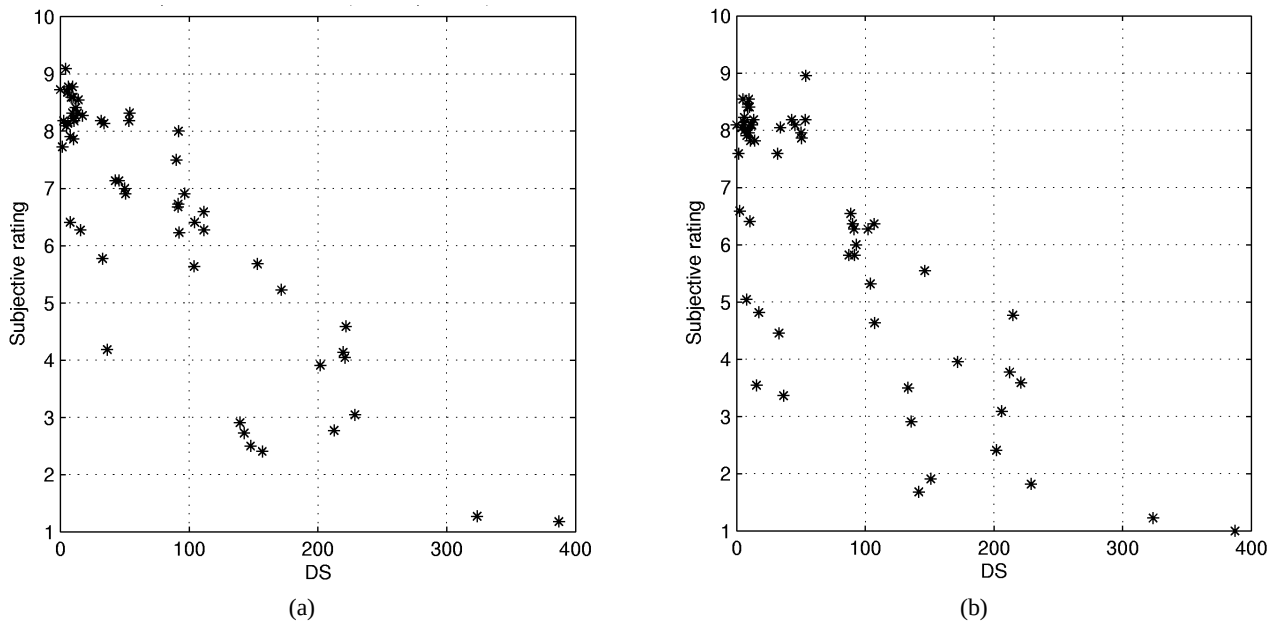

Fig. 11. As Fig. 10, but for ratings from experiment 2, which include distortions with pre- and postfiltering. (a) Speech test. (b) Music test.

a high degree of consistency of the group means across test sessions.

The pattern of results was reasonably consistent across subjects. Correlations between individual scores and mean scores, calculated as for experiment 1, are shown in Table 4. The correlations range from 0.75 to 0.94, indicating good consistency across subjects, although the consistency is not quite as high as for experiment 1. The SD of the ratings across subjects for a given condition was typically about 1.6 scale units (SE about 0.5 scale units), which is slightly higher than for experiment 1, but similar to experiment 2.

As before, ANOVAs were conducted on the ratings with subject and amounts of distortion as factors. This was done only for the artificial distortions. For the distortions involving prefiltering, the center frequency of the band where the distortion was introduced was also a factor. ANOVAs were calculated separately for the speech and music stimuli. The effect of the amount of distortion was always highly significant ($p < 0.001$). For the broad-band distortions, the effect of subject was sometimes significant, indicating that some subjects gave lower or higher overall ratings than others. The interaction of subject and value of α was significant for speech, $F(30, 44) = 1.99$, $p = 0.018$, and for music, $F(30, 44) = 1.97$, $p = 0.02$. However, in both cases the interaction accounted for less than 1% of the variance in the data. For the distortions involving prefiltering, the interaction of subject and value of α was significant for speech, $F(10, 44) = 5.59$, $p < 0.001$, and for music, $F(10, 44) = 6.48$, $p < 0.001$, but in both cases the interaction accounted for less than 2% of the variance in the data. We conclude, as before, that the ratings are mainly determined by the amount and type of distortion in the stimuli, and that individual differences in the patterns of ratings are small.

6.4 Comparison of Obtained Ratings and Distortion Score

As mentioned earlier, all stimuli in this experiment were bandpass filtered between 301 and 15 900 Hz (and spectrally shaped where appropriate), so as to remove as far as possible differences in overall spectral shape of the outputs for the differently distorted signals. This was done to allow us to isolate the perceptual effects of nonlinear distortion by equating linear distortion across conditions. In our previous analyses using the multitone test signal, that signal covered a wide frequency range, from 50 to 15 000 Hz. Such a signal would have been inappropriate to use here, since the input would have contained frequency components that were not present in the output of the systems under test. To avoid this problem, the input multitone signal was bandpass filtered between 301 and 15 900 Hz, in the same way as the stimuli were bandpass filtered in the experiment. Otherwise the calculation of the DS score was done in the same way as before.

Fig. 12 compares the mean ratings for the different systems with the DS score obtained using the multitone signal as input to the systems. There is a moderately strong relationship between the ratings and the DS values. The

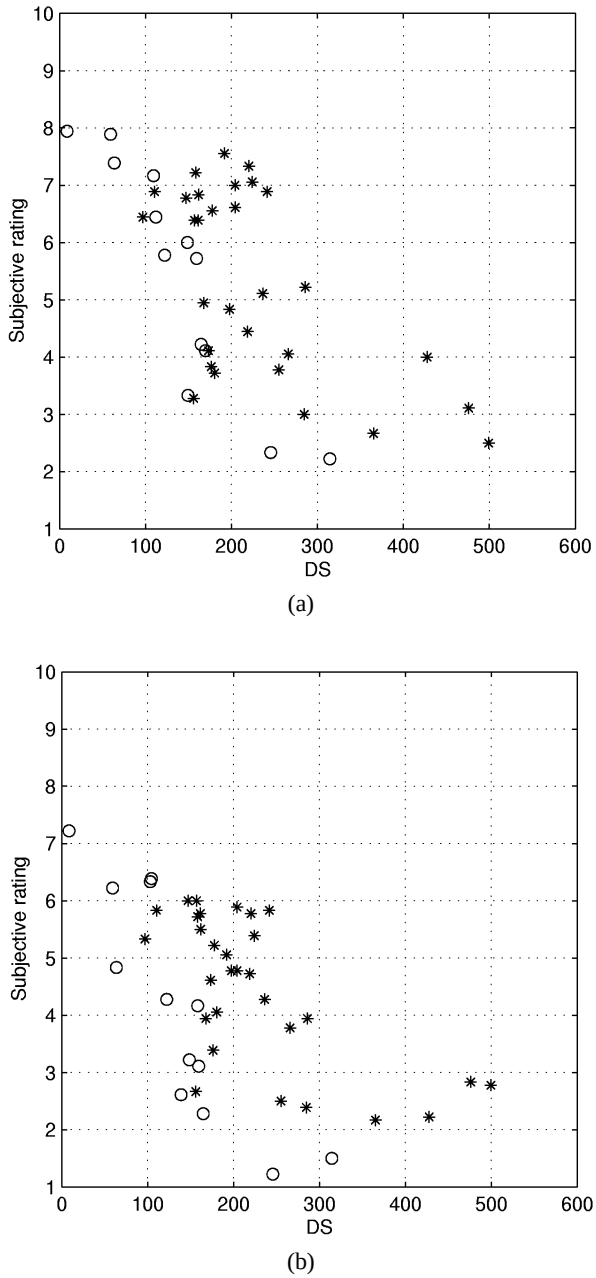

Fig. 12. As Fig. 10, but for ratings from experiment 3, which include distortions introduced by real transducers. (a) Music test. (b) speech test. ○ Artificial distortions; ✻ real distortions.

Table 4. Correlation of mean ratings across sessions for each individual subject with mean ratings across subjects, for experiment 31.

Subject	1	2	3	4	5	6	7	8	9
Speech	0.83	0.85	0.88	0.88	0.80	0.83	0.92	0.93	0.85
Music	0.77	0.94	0.90	0.84	0.91	0.75	0.81	0.92	0.86

correlations is -0.64 ($p < 0.001$) for speech and -0.595 ($p < 0.001$) for music. It is clear, however, that there is rather a lot of scatter. It appears that the relationship of the subjective ratings to the DS scores is somewhat different for the artificial distortions (circles) and the real distortions (asterisks). This was confirmed by calculating the correlation between the ratings and the DS scores separately for the artificial and real distortions. The correlations for the artificial distortions were -0.91 ($p < 0.001$) for speech and -0.86 ($p < 0.005$) for music. The correlations for real distortions were -0.60 ($p < 0.005$) for speech and -0.67 ($p < 0.005$) for music. Evidently the DS measure is more closely related to the subjective ratings for the well-defined artificial distortions than for the more complex real distortions.

It is noteworthy that the mean scores for the undistorted stimuli were relatively high, at 7.94 and 7.22 for speech and music, respectively, even though those stimuli were bandpass filtered between 301 and 15 900 Hz. In our previous study of the perceptual effects of linear distortion [10] we found that high-pass filtering at 313 Hz led to mean ratings of 2.1 for speech and 2.2 for music. The large difference in ratings across the two experiments indicates that subjects adjust their criteria depending on the range and type of stimuli presented. In experiment 3 all stimuli lacked frequency components below 301 Hz, and hence sounded rather tinny. However, the stimuli without nonlinear distortion clearly sounded better than the stimuli with strong nonlinear distortions, and the ratings given by the subjects reflected this. It is likely that the example stimuli played before the experiment proper contributed to the adjustment of the criteria of the subjects, and encouraged them to use much of the available response range. Although subjects clearly did change their criteria across experiments, the ratings given to the bandpass filtered undistorted stimuli in experiment 3 were lower than the ratings given to the broad-band stimuli with little or no distortion in experiment 1 and 2, which were typically about 9.

7 SUMMARY AND CONCLUSIONS

We have examined the effects of various types of nonlinear distortions on the perceived quality of speech and music signals. Subjects were presented with a series of distorted signals, with the different types of distortion presented in a randomized order, and were required to rate the perceived amount of distortion on a scale from 1 to 10, where 1 corresponds to most distorted and 10 corresponds to least distorted. The results for all experiments were highly consistent across subjects and test sessions. The ratings are largely determined by the amount and nature of the distortion in the signals, individual differences making only a small contribution to the variance in the ratings. Center clipping and soft clipping had only small effects on the ratings, whereas hard clipping and the full-range distortions had large effects. When the frequency components introduced by the distortion were limited to a specific frequency range (defined in ERB_N), all of the four ranges tested were approximately equal in importance in affecting perceived distortion.

The subjective ratings obtained from the first two experiments, which used various artificial forms of distortion, were compared to physical measures of distortion based on multitone test signals. A distortion measure DS derived from the output spectrum of each nonlinear system in response to a 10-component multitone signal gave high correlations with the subjective ratings. The correlations were negative because high ratings were associated with low values of DS. For experiment 1 the correlations were -0.98 and -0.97 for speech and music, respectively. For experiment 2 the correlations were -0.87 and -0.80 for speech and music, respectively.

A further experiment was conducted using stimuli for which nonlinear distortion was introduced by recording the outputs of real transducers. The output signals were digitally filtered to reduce irregularities in amplitude–frequency response as far as possible. The results showed reasonably strong negative correlations between the subjective ratings and the objective measure DS; for the stimuli recorded through real transducers, the correlations were -0.60 and -0.67 for speech and music, respectively. However, the correlations were not as high as for the artificial distortions. We conclude that an objective measure of nonlinear distortion based on the use of a multitone signal can predict reasonably well the perceptual effects of nonlinear distortion.

8 ACKNOWLEDGEMENT

This work was supported by Nokia Research Center (Finland). The authors wish to thank Kalle Koivuniemi (Nokia Research Center) for performing the acoustical measurements of the real transducers. They also thank Tom Baer, Brian Glasberg, and Michael Stone for their assistance with various aspects of this work, as well as Rosalie Uchanski and two anonymous reviewers for helpful comments on an earlier version of this paper.

9 REFERENCES

[1] R. Plomp, "Timbre as a Multidimensional Attribute of Complex Tones," in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G. F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands, 1970).

[2] R. Plomp, *Aspects of Tone Sensation* (Academic, London, 1976).

[3] A. Gabrielsson, B. N. Schenkman, and B. Hagerman, "The Effects of Different Frequency Responses on Sound Quality Judgments and Speech Intelligibility," *J. Speech Hear. Res.*, vol. 31, pp. 166–177 (1988).

[4] A. Gabrielsson, B. Hagerman, T. Bech-Kristensen, and G. Lundberg, "Perceived Sound Quality of Reproductions with Different Frequency Responses and Sound Levels," *J. Acoust. Soc. Am.*, vol. 88, pp. 1359–1366 (1990).

[5] A. Gabrielsson, B. Lindström, and O. Till, "Loudspeaker Frequency Response and Perceived Sound Quality," *J. Acout. Soc. Am.*, vol. 90, pp.707–719 (1991).

[6] B. C. J. Moore, *An Introduction to the Psycho-*

logy of Hearing, 5th ed. (Academic, San Diego, CA, 2003).

[7] A. Gabrielsson, P. O. Nyberg, H. Sjögren, and L. Svensson, "Detection of Amplitude Distortion by Normal Hearing and Hearing Impaired Subjects," *Karolinska Institute, Tech. Audiology*, vol. TA 83, pp. 1-20 (1976).

[8] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone Testing of Sound System Components—Some Results and Conclusions, Part 1: History and Theory," *J. Audio Eng. Soc*., vol 49, pp. 1011–1042 (2001 Nov.).

[9] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone Testing of Sound System Components—Some Results and Conclusions, Part 2: Modeling and Application," *J. Audio Eng. Soc*., vol 49, pp. 1181–1192 (2001 Dec.).

[10] B. C. J. Moore and C. T. Tan, "Perceived Naturalness of Spectrally Distorted Speech and Music," *J. Acoust. Soc. Am*., vol. 114, pp. 408–419 (2003).

[11] ANSA S3.22-1996, "Specification of Hearing Aid Characteristics," American National Standards Institute, New York (1996).

[12] J. M. Risch, "A New Class of In-Band Multitone Test Signals," presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 46, p. 1037 (1998 Nov.), preprint 4803.

[13] J. M. Kates, "A Test Suite for Hearing Aid Evaluation," *J. Rehab. Res. Devel*., vol. 27, pp. 255–278 (1990).

[14] D. A. Preves, "Expressing Hearing Aid Noise and Distortion with Coherence Measurements," *Asha*, vol. 32, pp. 56–59 (1990).

[15] O. Dyrlund, "Coherence Measurements in Hearing Instruments, Using Different Broad-Band Signals," *Scand. Audiol*., vol. 21, pp. 73–78 (1992).

[16] J. M. Kates, "On Using Coherence to Measure Distortion in Hearing Aids," *J. Acoust. Soc. Am*., vol. 91, pp. 2236–2244 (1992).

[17] J. M. Kates and L. Kozma-Spytek, "Quality Ratings for Frequency-Shaped Peak-Clipped Speech," *J. Acoust. Soc. Am.*, vol. 95, pp. 3586–3594 (1994).

[18] J. M. Kates, "Cross-Correlation Procedures for Measuring Noise and Distortion in AGC Hearing Aids," *J. Acout. Soc. Am*., vol. 107, pp. 3407–3414 (2000).

[19] ITU-R BS. 1387.1, "Method for Objective Measurements of Perceived Audio Quality," International Telecommunications Union, Geneva, Switzerland, pp. 1–100.

[20] ITU-T P.862, "Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Codecs," International Telecommunications Union, Geneva, Switzerland, pp. 1–24.

[21] M. D. Burkhard and R. M. Sachs, "Anthropometric Manikin for Acoustic Research," *J. Acoust. Soc. Am*., vol. 58, pp. 214–222 (1975).

[22] E. A. G. Shaw, "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1848–1861 (1974).

[23] G. Kuhn, "The Pressure Transformation from a Diffuse Field to the External Ear and to the Body and Head Surface," *J. Acoust. Soc. Am*., vol. 65, pp. 991–1000 (1979).

[24] M. C. Killion, E. H. Berger, and R. A. Nuss, "Diffuse Field Response of the Ear," *J. Acoust. Soc. Am*., vol. 81, p. S75 (1987).

[25] J. E. Jacobs and P. Wittman, "Psychoacoustics, the Determining Factor in Stereo Disc Distortion," *J. Acoust. Soc. Am.*, vol. 12, pp. 115–123 (1964).

[26] T. Letowski, "Difference Limen for Nonlinear Distortion in Sine Signals and Musical Sounds," *Acustica*, vol. 34, pp. 106–110 (1975).

[27] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc*., vol. 45, pp. 224–240 (1997 Apr.).

[28] E. Larsen and R. M. Aarts, "Reproducing Low-Pitched Signals through Small Loudspeakers," *J. Audio Eng. Soc*., vol. 50, pp. 147–164 (2002 Mar.).

[29] E. C. Poulton, "Models for the Biases in Judging Sensory Magnitude," *Psych. Bull.*, vol. 86, pp. 777–803 (1979).

[30] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hear. Res*., vol. 47, pp. 103–138 (1990).

[31] E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *J. Acoust. Soc. Am*., vol. 33, p. 248 (1961).

THE AUTHORS

C. T. Tan

B. C. J. Moore

N. Zacharov

Chin-Tuan Tan received B.E., M.E., and Ph.D. degrees in electrical and electronic engineering from the Nanyang Technological University (Singapore) in 1992, 1996, and 2000, respectively. His doctorate was on "Zero-Crossing-Based Compression Hearing Aids." He then joined Brian Moore's Hearing Group and now works as a postdoctoral research associate in the Department of Experimental Psychology, University of Cambridge. His present research project focuses on the perception of distortion in speech and music. His research interests include speech and auditory processing, digital signal processing, and hearing aids. He plays basketball and sings in the Wolfson College Choir.

●

Brian Moore is a professor of auditory perception at the University of Cambridge, UK. He is a fellow of the Royal Society, the Academy of Medical Sciences, and the Acoustical Society of America; and an honorary fellow of the Belgian Society of Audiology and the British Society of Hearing Aid Audiologists. He is president of the Association of Independent Hearing Healthcare Professionals (UK). He is a member of the Editorial Boards of Hearing Research, The International Journal of Audiology, and Audiology and Neuro-Otology. He is a consultant for several USA and European companies. He has written or edited 12 books and over 370 scientific papers and book chapters. He was recently awarded the Acoustical Society of America's Silver Medal in physiological and psychological acoustics. He is wine steward of Wolfson College, Cambridge.

●

Nick Zacharov was born in London in 1969. He obtained a Bachelor degree in Electroacoustics from Salford University, UK, and Master of Science and Doctor of Science degrees in technology from the Helsinki University of Technology, Acoustics and Audio Signal Processing in 1997 and 2002, respectively. After completing a one-year training period while in the UK with the Finnish loudspeaker manufacturer, Genelec, Dr. Zacharov returned to Finland as a design engineer with Genelec for over two years. He is currently working for Nokia Research Center in Tampere, Finland, as a principal scientist in the field of audio quality assessment. He is also a chartered engineer. His present research interests include spatial sound reproduction systems, audio quality, and perceptual evaluation methods.

Dr. Zacharov is a member of the Institute of Acoustics and the Acoustical Society of America and a vice-chairman of the AES Finnish Section. He was also the vice-chairman of the AES 16th International Conference on Spatial Sound Reproduction, 1999; co-chairman of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, 2002; and an AES governor. He has a number of publications and patents in the field of audio. He enjoys photography, cookery, downhill skiing, travel, and audio.