SOUND QUALITY MEASUREMENTS OF AUDIO SYSTEMS BASED ON MODELS OF AUDITORY PERCEPTION

Matti Karjalainen

Helsinki University of Technology
Acoustics Lab., Otakaari 5 A
02150 Espoo 15, Finland

ABSTRACT

Traditional methods of analyzing the sound quality of audio equipment are based on simple test signals as well as on simple measurement procedures. Sine wave or sweep is applied when measuring nonlinear (harmonic) distortion or frequency response. Combinations of sine waves and/or some kind of noise signals are the most complex test signals used in practice. Analysis methods as well are based on oversimplified models of auditory perception. It is easy to demonstrate how measures like harmonic distortion may be misleading when used to describe the perceived quality of sound. A new methodology is introduced here, which is based on a computational model of auditory perception as the analyzer of sound quality. The model makes it possible to use any real signal as a test signal. Some measurements carried out in practice are described.

AUDITORY SPECTRUM ANALYSIS

The human auditory system may be seen as a spectrum analyzer which differs from simple technical analyzers in many ways. The physiology of cochlear mechanics and the psychoacoustic theory imply a special inherent representation of auditory spectrum. The most important properties, when compared to simple Fourier analysis, are:

- spectral emphasis by the inverse of the equal loudness curves,
- so called Bark scale (critical-band scale) instead of Herz frequency scale,
- frequency domain resolution of about one critical band (1 Bark),
- masking effect in frequency domain and spreading of the spectral components, and
- masking effect in time domain (forward and backward masking).

All these properties are known from the traditional psychoacoustic theory /1/, but only few attempts are made to apply them to objective measurement of sound quality. Schroeder et al /2/ have used a computational model when evaluating signal-to-noise ratios in speech transmission.

Schroeder, Atal and Hall /3/ have given a mathematical model for auditory spectrum analysis. We adopted this model with minor modifications. The main phases of our analysis algorithm are:

- Fast Fourier transform (FFT) with 35 ms Hamming window.
- Emphasis of the spectrum by an approximation of the frequency sensitivity curve of the ear.
- Transformation of frequency f to Bark variable x by /3/:

  $x = 7 * \mathrm{arsinh}(f/650\mathrm{Hz})$

- So called "exitation function" E(x) is found by smoothing the Bark-scaled pre-emphasized power spectrum S(x) with a "spreading function" B(x):

  $E(x) = S(x) * B(x)$, (* means convolution)

  where B(x) is a piecewise approximation of the Schroeder et al spreading function /3/ by linear slopes (+25db/Bark, -10dB/Bark) and power series approximation for the top of the curve (see Fig. 1b).

- dB-scaled E(x) is the final auditory spectrum used in the study. Two examples of such spectra of simple signals are shown in Fig. 1. Auditory impulse spectrum (1a) has a form which is similar to the frequency sensitivity function of the ear. Auditory sine wave spectrum (1b) gives the form of the spreading function B(x).

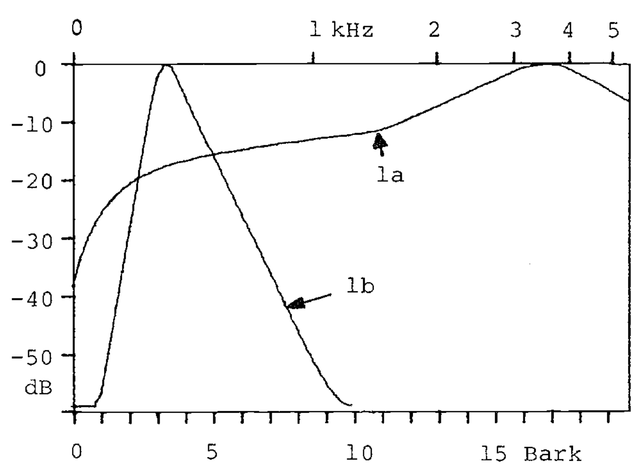

FIG. 1. Auditory spectra of simple test signals: impulse spectrum (a) showing the shape of frequency sensitivity of the ear; sine wave spectrum (b) corresponding to the form of the spreading function B(x).

AUDITORY SPECTRUM DISTANCE AS A MEASURE OF DISTORTION

If a sound is changed in some aspects, the perceived difference is found to correlate well with the corresponding change in auditory spectrum. From the theory of psychoacoustics we know that a deviation of 1 - 2 dB over any critical band (1 Bark) is the just noticeable difference (JND) found in listening tests /1/. We can expect that a proper measure of auditory spectrum distance between the original and the distorted signal will work as a natural measure of perceivable distortion. Fig. 2 shows the basic idea in block diagram form.

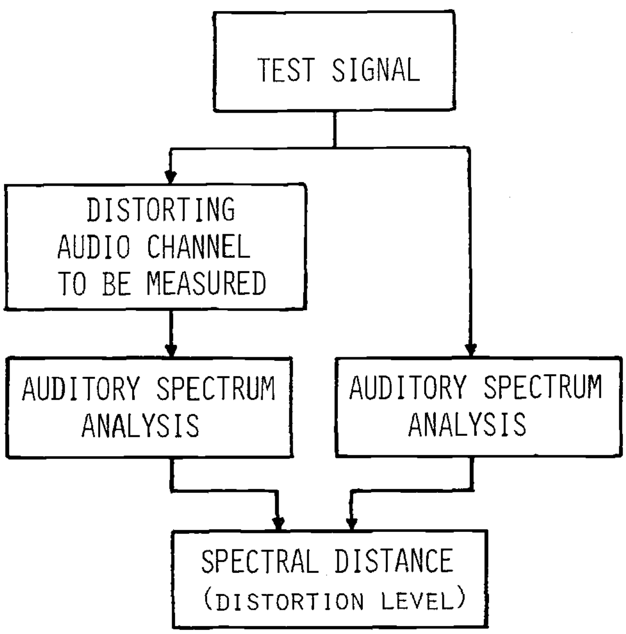

FIG. 2. The basic idea on the use of auditory spectrum distance as a measure of perceivable distortion.

There are different possibilities to define a function describing the auditory spectrum distance of two spectra. A simple but still useful measure is the maximum difference in any Bark-point between the E(x)/dB-spectra to be compared. (If difference in total loudness levels of the signals is not considered to be distortion, the levels should be matched according to power, loudness /1/, or average log level.)

Useful distance measures may be found by integrating the spectrum level difference $dE(x) = |E'(x)/dB - E(x)/dB|$ over the Bark scale:

$$D = \left(\int_x dE(x)^p \, dx\right)^{1/p} \qquad \text{(Plomp /4/)}$$

If p = 2 then D means the Euclidian distance of two auditory spectra E(x) and E'(x).

This way of measuring distortion by auditory spectrum distance will work without problems in cases where the envelope spectrum is changed due to any distortion but the periodic vs. noisy character, the fundamental frequency, or the existence of signal components is not essentially changed. The method as such can be applied both to linear (frequency response) and nonlinear distortion. If the test signal has harmonic spectrum or it is a random signal, nonlinear distortion also affects it so that the spectral distance works well. The general validity conditions of the method will be discussed later in this paper.

MEASUREMENTS OF AUDITORY FREQUENCY RESPONSE

As an example how to measure linear distortion (i.e. deviation from flat frequency response) we will discuss the case of loudspeaker measurements. Fig. 3 shows a series of auditory spectra. The test signal was a digitally recorded single hand clap (auditory spectrum in Fig. 3a).

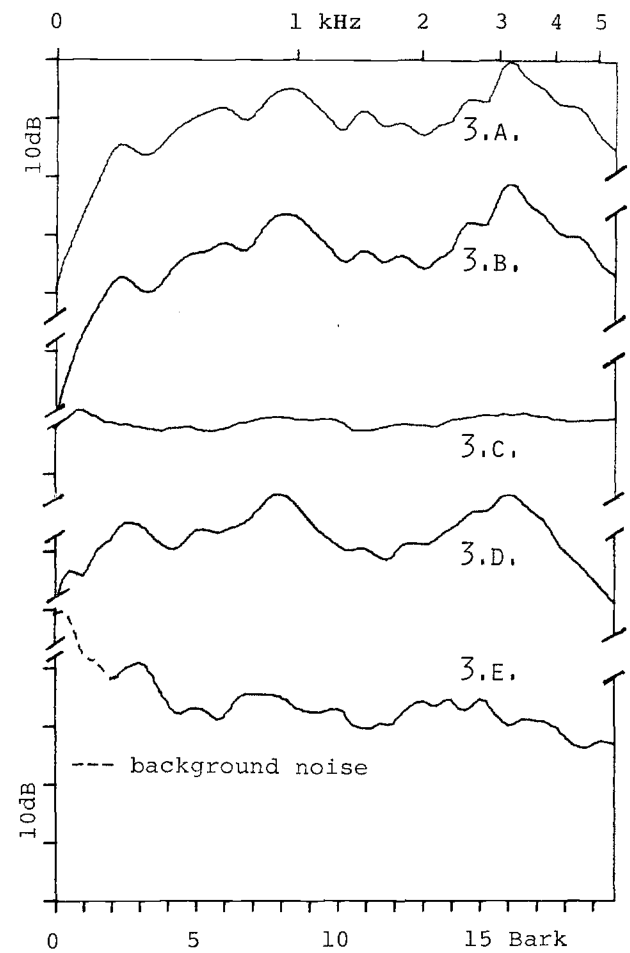

FIG. 3. Auditory spectra and frequency responses from loudspeaker measurements: (3a) auditory spectrum of the test signal (hand clap), (3b) loudspeaker response spectrum, (3c) auditory frequency response (AFR) of the loudspeaker, (3d) spectrum measured in a reverberant listening room from a sample 100 ms after the beginning of the test signal, and (3e) AFR of the loudspeaker-room-system corresponding to spectrum 3e.

The acoustic response of a small HiFi loudspeaker was measured and analyzed by using the model to get the auditory spectrum in Fig. 3b. Now we can define the auditory frequency response (AFR) to be the difference of spectra 3b and 3a. In our example this is shown in Fig. 3c. Very similar results can be found when using other test signals of fairly flat power spectrum. (The frequency scale was limited to 6.3 kHz because of limitations of the equipment.)

The method can be applied to the measurement of loudspeaker response in a reverberant listening room, too. By moving an analysis window in time we can see the AFR from direct response and first reflections to room reverberation. Auditory properties of the room acoustics can be evaluated in a similar way when using a known sound source. To be more accurate, a full dynamic auditory model with temporal masking properties included should be used instead of the present steady-state model.

MEASUREMENTS OF NONLINEAR DISTORTION

Nonlinear memoryless (static) distortion has the effect on the spectrum of a periodic or random signal of changing the relative amplitudes of the signal components. This depends on the properties of the signal and on the distorting mechanism. Also here the auditory spectrum distance is the natural objective measure of perceived degree of distortion. Fig. 4 illustrates how the auditory spectrum is changed when a speech sound (vowel /a/) is distorted in different degrees. The JND-level of distortion can again be assumed to correspond to about 2 dB of maximum deviation of auditory spectra, independently of the type of distortion and test signal. The most clearly pronounced effect of nonlinear distortion can usually be seen in the valley regions of non-flat spectra.

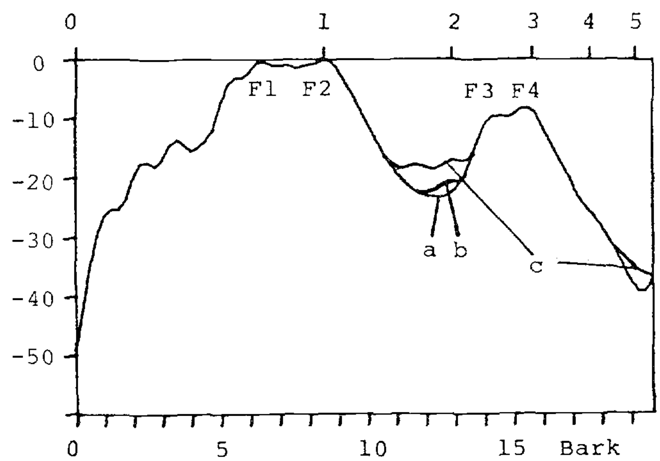

FIG. 4. The effect of nonlinear distortion on auditory spectra of vovel /a/: (a) no distortion, (b) just noticeable level of distortion, and (c) clearly noticeable distortion.

If the audio signal channel or device to be measured generates both linear and nonlinear distortion, they can be separated by first measuring the frequency response with a broad-band test signal and then the nonlinear distortion after compensating away the effect of the linear one.

EXPERIMENTS TO TEST THE VALIDITY OF THE APPROACH

Some experimental studies were carried out to show the validity and applicability of the approach /5/,/6/. The basic criterium was to compare the JND tresholds of distortion found by subjective tests to the values predicted by the auditory model. The main aim in our study has been the analysis of nonlinear distortion in speech signals /7/.

In the first study we used four Finnish vowel sounds /a,ä,i,u/ spoken by a male speaker. The signals were low-pass filtered (6.3 kHz because of equipment limitations) and computationally distorted by using a number of different nonlinear memoryless characteristics including quadratic, cubic, crossover, clipping, and other distortion mechanisms. The JND level of each type of distortion in each vowel was found by two trained listeners in direct comparison with undistorted signal. The model was then applied to compute the corresponding coefficients of distortion to give a maximum auditory spectrum distance of 2 dB.

The average difference between the subjective and the model-based results was only 2 dB so that the subjective detection levels were lower. Average differences for some distortion types were; crossover 5.5, clipping 0.1, quadratic 0.8 and cubic 1.4 dB. For all vowels and distortion types the range of differences was 6.3 dB. A still better match could be achieved by carefully trimming the computational model.

In another and more general experiment we tested the approach in the detection of any spectral change by using a signal source of periodic or random waveforms and with controllable spectrum shape. In fact, it was a parallel type speech synthesizer with five formant resonances. The JND-tresholds caused by any spectral changes (linear, nonlinear or parametric with resonance changes) and found in listening tests by one well experienced person corresponded to 1.5 - 3.1 dB Euclidean distance based on the model. This result shows that the perceptibility of any spectral change (not affecting pitch, fundamental frequecy, or degree of voicing/unvoicing) can be evaluated accurately and reliably by using the computational model. Auditory spectrum distance is a remarkably better correlate of subjective distortion perception than the traditional measure of harmonic distortion.

DISCUSSION

The proposed method of measuring linear distortion (AFR) is not very far from conventional measurement principles exept in its application of explicit model of auditory perception and practically any kind of test signal. Nonlinear distortion measurement method, we believe, has new and more fundamental aspects which make it to appear very similar to frequency response measurement. One can ask now in the light of the model, why should we take them apart, why not to use only one measure of auditory spectrum distance caused by any distortion mechanism.

The question can be answered that we have one measure only in a special case; when we deal with the JND treshold of distortion. Be it linear or nonlinear, the treshold values (direct comparison) can be predicted reliably by using the computational model. This special case is important from engineering point of view; it is the goal of quality design. No further improvement is audible. To be sure, we should find the worst practical test signal conditions, which may be a difficult task.

When the distortion level exceeds the JND treshold, the problem becomes much more complicated. Nonlinear distortion is easier to be perceived and more disturbing than the linear one because it is dependent on the signal and changing in time. Even a fairly high degree of constant response error may not be disturbing. Up to now we have investigated only the treshold level effects. Some subjects for further studies are:

- JND treshold of distortion detection (linear/nonlinear) when no undistorted reference is available.
- Nonlinear distortion detection for nonharmonic sounds (intermodulation effect).
- Distortion measures (over the JND treshold) from different points of view; "loudness" and disturbance of distortion with different signals, understandability of speech e.g. as a "phonetic distance measure", etc.

Besides the many other details we could discuss, there is one major point; how to include not only steady-state but also dynamic (temporal) properties of auditory perception to the models. This means two somewhat different aspects. The first one is to compute a short-term auditory spectrum as a function of time including forward and backward masking properties of the ear and the temporal integration of Bark-scaled power density with a time constant of 100 to 200 ms. We are developing a filter-bank model which is promising and seems to be accurate enough for practical purposes. It gives much more freedom to be used with short transient-like test signals than the present steady-state model.

Another important aspect is the analysis of the fine stucture of periodicity (vs. nonperiodicity) and how it is distorted in audio equipment. This seems to be quite different and maybe much more difficult question than the auditory spectrum. Decomposition of complex sounds into partials according to different sound sources and some kind of "auditory phase spectrum" should also be realized as explicit computational models.

One more subject to be commented here is the implementation possibilities of the auditory models. In our first study we used a microprocessor-based signal analysis system. It takes more than one minute to compute one auditory spectrum. Our dynamic models run in an array processor (FPS 100) but not in real time. New signal processors like NEC 7720 and TMS 320 will make it possible to implement very powerful steady-state models and also fairly practical dynamic filter-bank models. Within few years there could be practical measurement systems based on the approach.

SUMMARY

A new approach to the measurement of sound quality in audio systems is introduced which is based on computational models of auditory perception. The ultimate aim of the approach is to develop a methodology of objective measurements with good correlation to the results of subjective evaluations. This can, we believe, be achieved only by developing better models of human auditory perception. The paper summarizes our first results of applying auditory spectrum analysis to the mesurement of both linear and nonlinear distortion. The results are promising and have led us to the development of more complete dynamic models.

ACKNOWLEDGEMENTS

The work reported here is closely related to projects financed by the Finnish PTT and the Academy of Finland.

REFERENCES

/1/ R.Zwicker, R.Feldtkeller, "Das Ohr als Nachricten-empfänger". S. Hirzel Verlag, Stuttgart 1967.

/2/ M.R.Schroeder & al., "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear". 96th meeting of ASA, Honolulu Hawaii, 1978.

/3/ M.R.Schroeder, B.S.Atal, J.L.Hall, "Objective Measure of Certain Speech Signal Degradations Based on Masking Properties of Human Auditory Perception". In: Frontiers of Speech Communication Research (ed. Lindblom & Öhman), Academic Press 1979.

/4/ R.Plomp, "Timbre as a Multidimensional Attribute of Complex Tones". In: Freq. Anal. and Periodicity Detection in Hearing (ed. Plomp & Smoorenburg), Sijthoff, Leiden 1970.

/5/ M.Karjalainen, M.Virtanen, "The Measurement of Distortion in Telephone Sets" (In Finnish). Helsinki Univ. of Tech., Acoustics Lab., Report no. 26, 1981.

/6/ M.Karjalainen, "Measurement of Distortion in an Audio Signal Channel Based on Psychoacoustic Models". Proc. of NAS-82, Stockholm 1982, pp. 141 - 144.

/7/ M.Karjalainen, "Objective Measurement of Distortion in Speech Signals by Computational Models of Speech Perception". Proc. of 11th Int. Congr. on Acoustics (ICA-83), Paris 1983.