Transfer Function and Subjective Quality of Headphones: Part 2, Subjective Quality Evaluations

Françoise Briolle

C.E.R.D.S.M.
Le Brusc, Six-Fours 83140, France

and

Thierry Voinier

13010 Marseille, France

In order to evaluate the subjective quality of different headphones, a real time simulation of transfer functions is achieved by adding filters to a reference headphone. Listening with quick switching from headphone to another one becomes possible, then subjective quality evaluation method is improved. The results obtained with this method are compared with those obtained with real headphones.

0 INTRODUCTION

Listeners' subjective quality assessments are becoming more and more commonly used to evaluate electroacoustical equipment rather than objectively mesuring selected parameters. It is still difficult however to link up the objective acoustical characteristics of a given device with the subjective preferences.

It has therefore become necessary to improve the methodology of listening tests. An ideal listening test should produce results that are reproducible at any time and should reflect only the audibly perceptible characteristics of the electroacoustic equipment.

Many studies have been carried out on the subjective evaluation of loudspeakers [1], [2], [3], [4] and the listening tests have been quite well standardized [5]. This is not so in the case of headphones, on which only a few studies have been carried out [6], [7], [8].

A magnitude estimation method is proposed here for assessing and comparing the subjective quality of several headphones. In order to improve the conventional method and to be sure that only the audibly perceptible characteristics of the headphone are taken into account in making the subjective estimations, a new experiment was designed : the acoustical characteristics of the basic headphone were modified by means of filters in order to simulate various other headphones. It is possible in this way to listen with a single headphone to the sound produced by several headphone models.

1 EXPERIMENTAL METHOD

There exist several formal methods of measuring subjective responses.

The pair comparison method consists of asking the listener "which is better than" or "worse than" the other. With this method one can obtain a simple rank ordering which does not reflect the distances between the subjective impressions.

With scaling methods [9], the listener is given a scale, which is often a verbal one ranging from excellent to fair or bad. When the scale is a numerical one, the listener is usually required to rate his impresssions between 0 and 10.

A more direct method is "magnitude estimation" [10], involving no imposed scale, in which listeners are required to estimate the strength of an event as a proportion of its original or reference intensity. This method, which has also been used in sociology and in psychology (to make esthetic assessments), was used first by Stevens to mesure the relationship between loudness and sound intensity.

In the first part of the present study, the subjective quality of twelve different headphones was evaluated by twenty listeners, using the magnitude estimation method [11].

Each listener was asked to rate the quality of the headphones successively by giving a number proportional to the quality of the sound. The subjects could choose any numerical scale since no fixed scale was imposed. The tests were all carried out with the same musical passage lasting 1 mn (Glenn Miller. "In the Mood"). The headphones were presented in the same order to all the listeners.

The experiment was repeated a few days later under the same conditions. except that the order of presentation of the headphones was different.

The results presented in figure 1 are the averages of the twenty listeners' ratings in the first and second experiments.

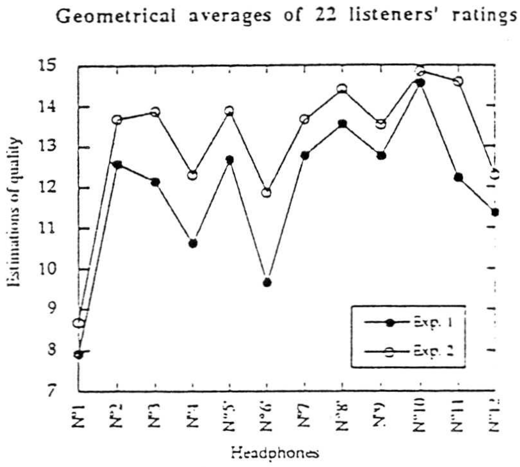

Figure 1

The results were consistent since the correlation between the two sets of ratings was equal to 0.93.

The magnitude estimation method is therefore a suitable means of evaluating the subjective quality of headphones.

In order to improve the conventional method however we developed a modified version of the above experimental procedure with which it is possible to listen on a single headphone to the sound produced with several headphone models [12]. Here the acoustical characteristics of the basic headphone (named the simulator headphone). were modified by means of filters in order to simulate various headphones.

2 EXPERIMENTAL DESIGN

2. 1 Experimental set-up

The signals reaching the listener's ears are filtered by the headphone. If we assume a headphone to be a linear system. its acoustical characteristics will be given by its transfer function (or its impulse response). The input signal is convolved by the impulse response of the headphone [13].

Due to the filtering. the simulator headphone will have the same acoustical characteristics as any given headphone. To perform the simulation. the transfer function of the filter is taken to be the ratio of the transfer function of the headphone to the transfer function of the simulator headphone[14] (for further details. see part 1).

In order to obtain a stereophonic input signal. two filters are needed. one for each channel. to simulate a given headphone.

The filtering is achieved by means of two DSP boards (Motorola 56001) connected to Macintosh II : the signal can be convolved with the impulse responses of the parallel filters (200 taps) in real time at a sampling frequency of 44.1 kHz.

In the present experiment. filters were computed for use with several simulators in order to simulate the acoustical characteristics of various headphones .

With this experimental set-up. the listener wearing a single simulator headphone can listen in turn to several headphones simulated by the various filters and assess their quality taking only their acoustical characteristics into account (with real headphones. other factors such as comfort. weight etc. might influence the listener's estimations). Switching from one simulated headphone to another. by simply changing the coefficient of the digital filters. takes less than 30ms: the testing time is thus considerably reduced. The whole experiment can be managed by the computer. which means that the experimental error is also reduced.

2.2 Experimenter interface

The interface with the experimenter was provided by HyperCard. Some external commands were added to HyperCard in order to run the DSP boards.

The experimenter could easily set up the experiment and select:

-the simulated headphones (filters) to be presented to the listener.

-the simulator headphone to be used for the experiment.

-the source of the sound (a CD player or a DAT).

The experiment was managed by the host computer. Each simulated headphone was chosen at random among those selected and the filtering was carried out by the DSP boards. At the end of the musical extract. the listener was asked to rate the simulated headphone quality. The listener's ratings were recorded.

2.3 Listener interface

The listener interface (VT320 terminal) was connected to the Macintosh II through the serial port. The subject was placed in a sound proof room and was informed about the experiment through the terminal screen and asked to use the numerical keypad to rate the quality of the headphone he was listening to.

3 LOUDNESS EQUALIZATIONS

If all the headphones to be compared do not have the same loudness. the listener might be disturbed and the ratings affected [15]. [16].

The loudness of all the simulated headphones was therefore adjusted. Two methods were used, namely subjective and objective equalization methods.

3.1 Subjective loudness equalization on a white noise

Subjective equalization of the simulated headphones was carried out by six listeners on a white noise, at a level of around 70 phones.

The listener was asked to adjust the level of all the simulated headphones to the same loudness, using a digital attenuator. He had to set a reference loudness by selecting one of the simulated headphones and then adjust the level of the other headphones to it, one by one, using a pair comparison method.

The listeners required on the average a dynamic range (max level -min level) of 9 dB to equalize the loudness of the twelve simulated headphones. The standard deviation of six listeners' adjustments was less than 1dB, with each simulated headphone. The various listener adjustments were in excellent agreement: the correlation coefficient between each of the listener's adjustments performed on the twelve simulated headphones and the average of all the listeners combined was greater than 0.90.

The consistency of the listeners' adjustments was mainly due to the experimental set-up : since switching from one headphone to another one takes very little time (only 30ms), the task of loudness equalization is not really very difficult.

3.2 Objective equalization on a white noise

Objective equalizations were computed in order to standardize the energy of the various output signals from all the simulated headphones. Several studies have been carried out on this topic [17], [18]. We used a simple method: with a white noise as input signal, we computed for each simulated headphone the power spectrum (Ei) of the output signal in the 0 to 15 kHz frequency range (the output signal was weighted by the equal loudness contour at 70 phones). If the lowest energy is E0, the attenuation of each headphone is equal to:

$$A_i = 10 \log (E_i/E_0)$$

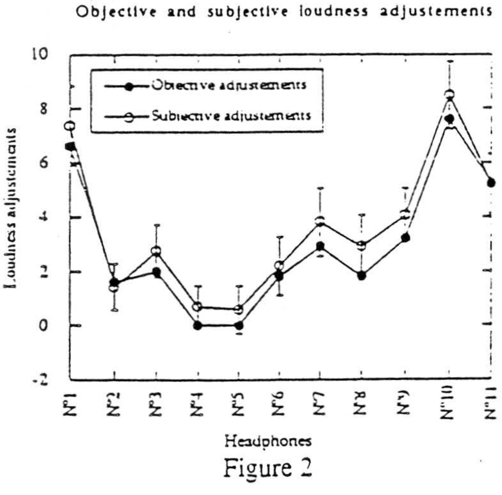

Figure 2

Figure 2 gives the objective and subjective adjustments with various simulated headphones.

They differed by less than 1 dB in the case of all the headphones. Similar results were obtained with loudspeakers [18].

3.3 Loudness equalization on various musical excerpts

The subjective loudness equalization procedure was repeated with two different pieces of music.

In the case of music, objective equalizations can be computed using a low pass filtered white noise (Fc=5kHz, 3dB/octave) as the input signal, in order to keep roughly within the spectrum of the pieces of music. Similar results were obtained to what was observed with a white noise : excellent agreement between various subjective adjustments (correlation coefficients greater than 0.90) and differences of less than 0.5 dB between objective and subjective adjustments with each simulated headphone.

With this experimental set-up, the loudness of the various simulated headphones can therefore be satisfactorily equalized either by performing objective calculations or by asking a few listeners to perform subjective equalizations.

4 INITIAL RESULTS OF SUBJECTIVE QUALITY ESTIMATIONS

4.1 Reliability of the quality estimations

The subjective quality of eleven different simulated headphones was evaluated by nine listeners, using the same extract from a spoken recording, lasting 15 s. The simulator headphone used for this experiment was headphone n°5 (STAX lamda Pro). The magnitude estimation method was used to assess the quality.

Each listener participated in six experiments, and therefore rated each headphone six times on various occasions. At each experiment the headphones were presented in random order.

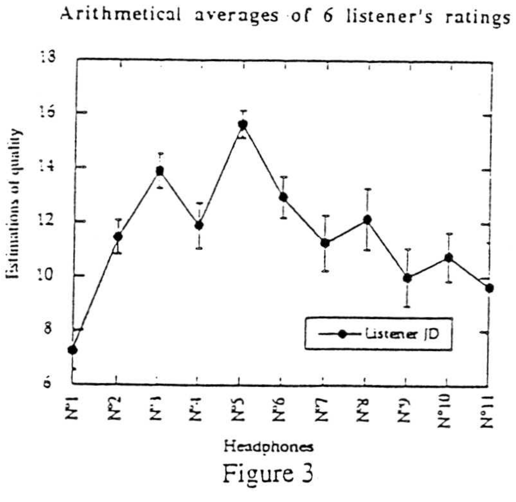

Figure 3

Figure 3 gives one listener's average rating and the standard deviation.

With each listener, the correlation coefficient (Bravais-Pearson) was computed between all the sets of ratings, two by two. This coefficient was always higher than 0.60 ($0.60<r<0.96$) which means that all the ratings produced by each listener were significantly correlated at a probability threshold of $p = .05$ [19].

In experiments with real headphones, the twenty listeners participated in two experiments. Here the correlation coefficient between two sets of estimations was greater than 0.50 with only twelve of the subjects (the correlation between the two sets of estimations was significant at a probability threshold of $p = .10$) and greater than 0.70 with only six of them (significant at a probability threshold of $p = .01$).

With the simulated headphones, the listeners' estimations of the various headphones' quality were more reliable, mainly because it took only a short time to switch from one headphone to another with this set-up (around 30ms with simulated headphones as compared to 5 or 10s with real headphones).

4.2 General agreement among the listeners

In these experiments, since no scale was imposed, the listeners used quite different rating scales ranging between 4 to 6 and 15 to 75.

In order to balance out the ratings, we normalized the data as suggested by Stevens [10]: each listener's individual ratings were divided by the geometrical mean of his ratings. The average of the combined listeners' ratings was the geometrical mean of the normalized data. In this way, the relative values were preserved and the preponderance of large numbers was conterbalanced.

Figure 4 gives one of the listeners' normalized ratings on the various headphones and the averages of the combined listeners' ratings.

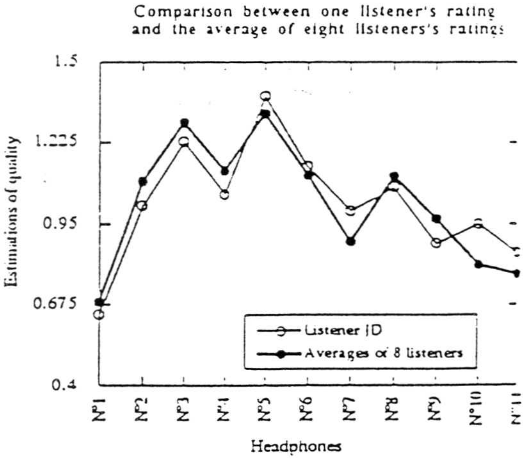

Figure 4

The correlation coefficient was calculated between each individual listener's ratings and the average of all the listeners' estimations. In eight cases, this coefficient was higher than 0.80 (which means that the correlation was significant at a probability threshold of $p = .003$) and in one case only, it was equal to 0.45 (which means that this listener's ratings of the various headphones were not correlated with the average combined listeners' ratings).

In order to compute the average of the other eight listeners' ratings and the standard deviation we used another type of normalization [20] which gives the advantage of normalizing the rating dynamic range used by the listeners. With each listener, the average of each set of normalized ratings was equal to 0 and the standard deviation equal to 1 (z transformation). The average of all the listeners' ratings combined was therefore the arithmetic mean of the normalized data.

Figure 5 gives the average of the other eight listeners' ratings and the standard deviation.

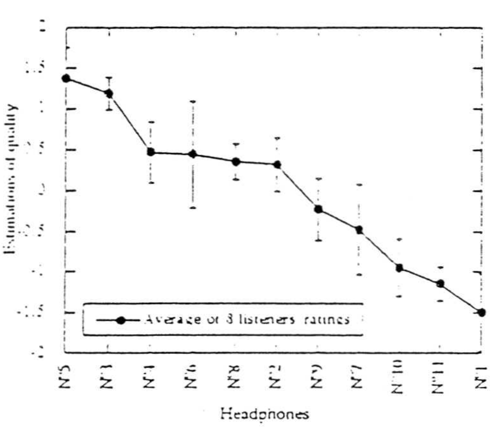

Figure 5

Allowing for a scaling factor, the results obtained after this normalization were very similar to those obtained with the normalization procedure recommended by Stevens.

4.3 Possible influence of the simulator headphone

Up to now, we have always used the headphone N°5 as simulator for our experiments. The acoustical characteristics of this headphone are really good (the frequency response is flat), so it is quite easy to modify the transfer function of this headphone in order to simulate others with less satisfactory acoustical characteristics. It is worth noting that headphone N°5 was rated by all the listeners as the best of all the simulated headphones.

Two experiments were carried out with two other simulator headphones. One of these was headphone N°1: its acoustical characteristics included large spectral alterations and it was rated by all the listeners as a poor quality headphone. The other one was headphone N°2, which was rated as a good headphone.

The subjective quality of the same eleven headphones, simulated by these two different headphones, was evaluated by six listeners using the same spoken passage.

The averages of all the listeners' normalised ratings with these three simulators are given in figure 6.

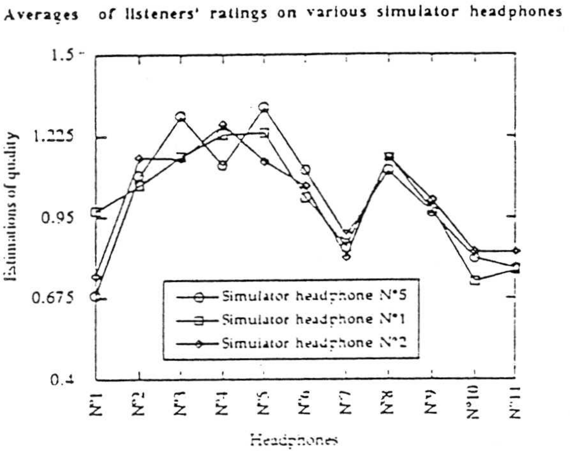

Figure 6

It can be seen from this figure that headphones N°5, N°1 and N°6 were always rated higher when they were simulated by themselves (i.e. when the impulse response of the simulator filters was a Dirac) than when they were simulated by another headphone. This is probably due to the precision of the filters: it was necessary to use 200 taps on-line for the impulse responses of the FIR filters (at a sampling frequency of 44.1 kHz) to be able to work in the real-time mode: otherwise the filters were not precise enough to simulate the acoustical characteristics of the real headphone.

The main point however is that the ratings obtained with different simulators were highly correlated (0.82<r<0.86 and 0.84<rs<0.87). In conclusion, it can be said that the listeners gave the various headphones the same quality ratings whichever headphone was used as the simulator.

5 CONCLUSIONS

An improved listening test procedure is decribed here whereby the acoustical characteristics of the basic headphone were modified by means of filters in order to simulate various other headphones. In this way, subjects are able to listen with a single headphone to the sound produced by several headphone models.

The simulation was carefully controlled from the acoustical point of view: the correlation coefficient between the measured transfer function of a real headphone and that of the simulated headphone was found to be higher than 0.85 (for futher details see part 1). This simulation method was also found to be satisfactory from the perceptual point of view: the ratings of the various simulated headphones were highly correlated whichever headphone was used as the simulator.

This experimental procedure is an improvement for several reasons: it considerably reduces the experimental time (in the case of headphones, switching from one headphone to another takes around 5 to 10s: with simulated headphones it takes less than 0.5s) and the experiment can be managed entirely by the computer, which makes for a lot of flexibility in randomizing the order of presentation of the stimuli and collecting and analyzing the data.

Furthermore, it is now possible to state with certainty that listeners' assessements reflect only the audibly perceptible characteristics of the electroacoustic equipment. In addition, we noted that the consistency of the listeners' assesements was improved, which makes it possible to link up the objective acoustical characteristics with the subjective preferences.

One most important results of this study is the finding that the quality of a headphone of poor or fair quality can be considerably improved simply by simulating on it the acoustical characteristics of an excellent headphone. This can be done in real time, with 200 taps FIR filters at a sampling frequency of 44.1 kHz.

In the case of headphones, the distortion is very small so that they can be modelled with a linear system and their acoustical characteristics can be said to be given by their transfer function. If this holds true in the case of other acoustical equipment such as loudspeakers, amplifiers and microphones, this simulation technique will also be applicable to these devices as a means of improving the listening tests as well as enhancing the quality of the actual equipment.

ACKNOWLEDGEMENTS

The authors are grateful to A. Freed for designing the DSP boards, and providing help for with the FIR programming. T. Angelini helped with the headphone and dummy head amplifiers design. Dr Jessica Blanc was kindly reviewed the manuscript.

6 REFERENCES

[1] Toole F.: "Subjective measurements of loudspeaker sound quality and listener performance" J.A.E.S., vol. 33, pp 2-32 (1985)

[2] Gabrielson A.: "Perceived sound quality of sound reproducing systems" J.A.S.A. 65(4), pp 1019-1033 (1979)

[3] Lipshitz S.: Vanderkooy J.: "The great debate: subjective evaluation" J.A.E.S., vol 29, n°7/8, pp 482-490 (1981)

[4] Toole F.: "Listening tests - Turning opinion into fact" J.A.E.S., vol 30, n°6, pp 431-445 (1983)

[5] International Electrotechnical Commission: "Sound System equipment. Part 13: Listening tests on loudspeakers" IEC report publication 268-13 (1985)

[6] Toole F.: "The acoustics and psychoacoustics of headphone" presented at the AES 2nd Int. Conference, Anaheim, CA, preprint C1006 (1984 May)

[7] Theile G.: "On the standardization of the frequency response of high-quality studio headphones" J.A.E.S. vol. 34, n° 12, pp 956-968 (1986)

[8] Lambert J.M.; Perdereau J. :"Evaluation de la qualité acoustique de casques d'écoute à l'aide d'un manequin équipé d'une oreille artificielle" Revue d'Acoustique, n°69, pp 85-90 (1984)

[9] Jones F. :"Overview of the psychophysical scaling methods" E.C. Carette and M.P. Friedman, Eds, Handbook of Perception, vol 2, Ch 10, Academic Press, New York (1974)

[10] Stevens S.: "Psychophysics" Wiley, New York, (1975)

[11] Briolle F.: "Evaluation subjective de casques d'écoute" Rapport Technique G.E.R.D.S.M., n° 50693 (1986)

[12] Posselt Ch.; Briolle F.: "Simulation of headphones", presented at the 116° ASA Meeting, (nov. 1988)

[13] Oppenheim A.; Schafer R.: "Discrete-time signal processing" Prentice Hall (1985)

[14] Papoulis A.: "The Fourier Integral and its applications" McGraw-Hill, New York (1962)

[15] Illényi A. ; Korpassy P.: "Correlation between loudness and quality of stereophonic loudspeakers" Acustica, vol 49, pp 334-336 (1985)

[16] Gabrielson A.; Hagerman B.; Bench-Kristensen T.; Lundberg G.: "Perceived sound quality of reproduction with different frequency responses and sound level" J.A.S.A. 88(3), pp1359-1366 (1990)

[17] Zwicker E.; Fastl H.; Dallmyr C.: "Basic programm for calculating the loudness of sound from their 1/3 oct band spectra according to ISE 532B" Acustica 55(1), pp 63-68 (1984)

[18] Aarts R.: "A comparison of some loudness measurements for loudspeaker listening tests" J.A.E.S., vol 40, n°3, pp 142-146 (1992)

[19] Faverge: "Méthodes statistiques en psychologie appliquée" PUF (1972)

[20] Keppel G.: "Design and Analysis: A researcher's Handbook" Prentice Hall (1982)