Subjective Testing of Compression Drivers

Earl R. Geddes, GedLee LLC, Northville, MI 48167

Lidia W. Lee, Eastern Michigan University, Ypsilanti, MI 48197

and

Roberto Magalotti, B&C Speakers, Florence, Italy

Abstract

A subjective test was devised and performed in order to assess the factors that influence the perception of sound emitted by compression drivers. A musical passage was high-pass filtered and played through three compression drivers of similar characteristics, loaded by a plane wave tube, and recorded. To obtain different levels of nonlinear distortion, the passage was played at three different voltage levels on each driver. The resulting sound files were recombined with the low-pass filtered portion, yielding nine complete sound pieces whose only differences to the original passage were caused by the drivers' behavior. The nine stimuli were then presented, in a double-blind test, to 27 subjects, who were asked to rate audible differences when compared to the original passage. Analysis of the results shows that the differences in frequency response between drivers is statistically significant, while differences in playing level, and therefore nonlinear distortion, was not significant. This unexpected result implies that nonlinear distortion is not audible under these test conditions and leads to important conclusions regarding the design objectives of compression drivers.

0 Introduction

The background for this report is based on the desire to make design changes in compression drivers that will positively affect the perception of these products in the marketplace. The primary goal – to optimize a compression driver design for its subjective perception - requires two separate, but related, pieces of information. The first is the *subjective* factors that influence the perception of the product, and the second is, the *objective* technical factors that influence the significant subjective factors. Hence, two main tasks to effecting a positive change to the perceived quality of a compression driver are required. The first task needs to examine those aspects of perception that are the most significant in judgments of sound quality; and the second, to examine the objective factors of compression driver technology that influence the most significant perceptual aspects of its performance. Both tasks must be performed in order to obtain an efficient achievement of the primary goal. Simply knowing the technical aspects of a

compression driver's performance does not indicate how these factors influence the end user's perception of its sound quality, and simply knowing those factors that affect the perceived sound quality does not indicate what design changes are desirable. Thus, in addition to standard objective measurements, it is necessary to perform psychoacoustical experiments to determine how the technical aspects relate to the subjective perception. This is the central theme of this study.

This report focuses on the first (subjective) task as defined above. However, it should be noted that lacking either of these two pieces – objective and subjective - one can only guess at how the customer perception optimization is to be achieved.

1 The Experiment

An experiment was designed to determine the perceptual importance of two forms of distortion commonly assumed to exist in, and limit the performance of, a loudspeaker driver (specifically here, a compression driver); 1) linear distortion or frequency response; and 2) nonlinear distortion. In this experiment it is necessary to minimize extraneous and un-controlled variables, thereby leaving only the control variables - linear and nonlinear distortion (called *source* and *level* in this study, see Sec 3.1).

The preferred procedure for obtaining our first piece of information is to find those variables that are subjectively important with a *reference stimulus* test and subsequently, perform a quality-based assessment on those variables using a test protocol that offers the best potential for making a quality-based assessment (see appendix). In other words, first find those variables that are the most important and then quantify or scale these variables in greater detail.

2 Method

2.1 Experimental design

Two factors were examined utilizing a 3x3 (two factors with 3 conditions in each factor) experimental design. The first factor is the *source*, where three different compression drivers were used. The second factor is *level*, where signals were sent to each driver at three different voltage levels.

The drivers themselves were all high quality, current production models from three different manufacturers. They all have the same common characteristics as follows;

- titanium dome on 100 mm edgewound aluminum voice coil;
- five-pieces phasing plug (exiting on four annular slits);

- 50 mm throat;

- 8 ohm nominal impedance;

- NdFeB magnet

The actual frequency responses for the drivers must be withheld for confidentiality purposes. This information - which would certainly be of interest to the reader – is of no consequence to the results obtained in this experiment. Its disclosure would, however, expose the identities of the three drivers since frequency responses are a kind of signature of the products. They were all similar but do have some distinct differences particularly at the higher frequencies.

The recording levels were adjusted so that each stimulus had the same recorded *rms* level. As a result this would create stimuli which had virtually the same perceived loudness level on playback even though internally they were played at three different sound levels. This is true, of course, only because the stimuli were all derived from the same piece of music and had only small differences in the linear and nonlinear distortions which would not a significant effect on the perceived loudness levels.

A total of nine stimuli were used along the two dimensions of *source* and *level*. A statistically significant difference of the main effect *source* would indicate that the subjects could detect linear distortions (frequency response) in the drivers and a significant difference of the main effect *level* would indicate that the subjects could perceive nonlinear distortion in the drivers. Interaction effects between *source* and *level* are also possible and were analyzed.

A reference stimulus (the original unmodified passage) was available throughout the experiment. The subjects were encouraged to evaluate the level of perceptibility of distortion in the test stimulus when compared to the reference stimulus.

The stimuli were presented using the Eytmotic ER4b research insert earphones played back through a Turtle Beach Santa Cruz sound card. These earphones have a very low amount of internal distortion and a reasonable frequency response. Comparative listening by the authors did not find that the sound from the earphones was a factor in the experiment, as it would have been if the earphone distortion was greater than that which was in the actual stimulus itself. It is very unlikely that these earphones would exhibit the levels of distortion found in the drivers used in this test.

The experiment was performed double-blind, that is, the test administrator was unaware of the intention of the test or any of the variables being studied. Furthermore, as much information as possible was withheld from the test administrator to prevent any effect on the results that could occur should they

have known this information. The drivers will simply be referred to as *source* 1, 2, and 3 and the levels as *level* 1, 2 and 3.

2.2 Stimulus:

The musical passage that was selected for this experiment was a 15 second segment of <u>*Burning Down the House (live)*</u> by *Talking Heads*, starting 129 seconds into the recording. This piece was chosen because it was found in an earlier (unpublished) study that it led to a high sensitivity in the perception of signal distortion.

Only one channel of the original recording was used, in order to have a monophonic signal. This monophonic signal was passed through a third order Butterworth crossover network, having its crossover frequency at 800 Hz. Signal processing operations were made using the MatLab Signal Processing Toolbox and MathCAD.

The high-pass filtered signal and the low-pass filtered signal were recorded as the two channels of a PCM wave file. The signals were output and recorded by the sound card. The low-pass portion was fed back to one of the recording channels. The high-pass portion was amplified by a Crown Macro-Tech 5000VZ, fed to the compression driver under test and the acoustical output of the driver was recorded with a Bruel & Kjaer 6.3 mm pressure microphone, and sent to the other recording channel. The resulting recording then had the low-pass portion on one channel, as output from the sound card, and the high-pass portion on the other channel, as "played" by the compression driver under test.

The drivers were excited with the high-pass portion of the musical passage into a progressively damped plane wave tube [1]. Three voltage levels 14, 20 and 28 Vrms were tested (*level*). Figure (1) shows a typical driver (they were all similar) and the levels of total harmonic distortion products found at the three voltage levels. The high-pass filtered signal was equalized with a 6 dB/octave high frequency boost above 2 kHz implemented as a 21 tap FIR filter (Figure 2). This equalization, identical for each driver, is needed to compensate for the mass roll-off that occurs in compression drivers. With this equalization, the linear frequency response of all the compression drivers on the plane wave tube was reasonably flat, and reasonably uniform across drivers, in the useful frequency range.

Since the high-pass portion is delayed relative to the low-pass portion (due to the time of flight from driver to microphone), the final stimulus was created by normalizing the two portions and adding them back together after suitably delaying the low-pass one. The time delay and normalization levels were determined by sending noise through the low-pass channel and the compression driver on the plane wave

tube. By cross-correlating the input and output, the delay is found as the point of maximum correlation. Level matching of the noise signal is also quite straightforward by simply looking at the total spectrum of the recombined noise signal. A final tweaking of the delay due to different phase responses of the drivers was done to achieve the seamless transition of the signals through the crossover point.

As a result, the stimuli obtained contained only changes that are made within the compression drivers with all other parameters of the stimuli having been held constant. With this procedure, the only remaining differences between the nine stimuli were:

• small linear frequency response differences from driver to driver (*source* differences);

• nonlinear distortion content, within the same driver, according to the input voltage level (*level* differences).

It should be noted that there is no horn or waveguide placed on these devices as this was not the focus of the study. In a complete system the horn and the driver to horn matching would have a substantial effect on the perceived quality of the combination, but our goal was to determine the driver contribution to the sound quality of this system.

2.3 Subjects

Twenty-seven college students were recruited for the experiment. All subjects passed an audiometric screening test at 25 dB HL (ANSI, 1989) for 250, 500, 1000, 2000, and 4000 and 8000 Hz. All subjects were paid for their participation.

2.4 Test Protocol

A computer program was written in Visual Basic to facilitate the presentation of the stimuli, the recording of the data, and the tracking of the stability of the responses. First, the program presents a short training section where the subjects were given some contrived examples along with the suggested ratings, followed by the formal test.

During the formal test, each subject is presented with a stimulus selected at random and played from the beginning to the end (15 sec). After the complete presentation, the subject can give a rating or they can do a direct real time A-B comparison of the reference stimulus and the test stimulus. The results are recorded via a unitary sliding scale with 0 = "Imperceptible" to 1.0 = "Clearly Perceptible" with a label of "Barely Perceptible" (.5) in the middle. Ten steps between the extremes were possible for selection.

After rating, a new stimulus is presented and the program cycles through once again with a randomly selected stimulus.

After a set of nine test stimuli have been presented and rated, the complete set of nine stimuli are presented again, at random, two more times. After three trials for each stimulus have been completed, the data is checked to see if the responses were consistent. This is done by checking to see if all of the responses lie within the *acceptance range* from the mean of the three trials. The *acceptance range* is read by the program from an external data file. In this experiment it was set at .15 of the maximum scale. This means that a single data point that lays more than .15 away from the mean of the group is rejected and the trial rerun on that stimulus. This process continues until all of the trials are stable or the subject has tried, and failed, five times to give a consistent set of answers. This stability check typically rejects about 20% of the test subjects.

When a complete test has been performed the program writes out a data file with all of the individual responses followed by the statistical data for each stimulus.

3 Results

The ratings were analyzed in a two-way repeated-measures analysis of variance (3 *sources* by 3 *levels*) using a statistical program called SPSS. The results indicated a significant main effect for *source*, $[F(2,25) = 10.934, p<.001]$, however, no significant differences were observed for the main effect *level* $[F(2,25) = 1.7, p = .203]$, nor the interaction effect of *source* and *level* $[F(4,23) = .458, p = .765]$. This means that the subjects could detect a difference between the *sources* (the different compression drivers), but not the source playback *levels*. Thus, subjects could detect linear distortion differences, but not nonlinear distortion differences, which is a most interesting result.

Figure 3 shows the mean data across *level* and *source*. The X-axis represents the various *sources*, and the Y-axis represents the mean rating across subjects. The various *levels* are represented by the symbols (*level* 1- square, *level* 2 – diamond, *level* 3 – triangle). It can be seen that the variability within *level* is very small and that there is no consistency in these variations across the drivers (the variations for *level* are not statistically significant). The data indicates that *source* 2 has significantly more perceptible linear distortion when compared to the original sound segment. *Sources* 1 and 3 were virtually indistinguishable from each other and possibly indistinguishable from the reference (see section General Observations).

4 General Observations

Subjects will seldom use the extremes of the allowed number scale. Thus a value near zero or unity should not be expected. This means that Sources 1 and 3 may well have been "imperceptible". There is no doubt that Source 2 has perceptible linear distortion and significantly more than Sources 1 and 3. In hindsight it is apparent that the addition of a dummy source, one that was in fact the reference, would have added value to the interpretation of the results. It is most unfortunate that this possibility was not seen beforehand.

5 Summary and Conclusions

These results are important as well as interesting since the insignificance of Level was not anticipated. The fact that nonlinear distortion in a compression driver is not a significant subjective parameter is quite enlightening and useful. Prior to this study, the exact opposite was believed to be true; that nonlinear distortion was a major problem in compression drivers. Regardless of any previously held notions about the auditory importance of nonlinear distortion in compression drivers, what the data is saying must be accepted - nonlinear distortion in a compression driver is simply not a factor to its sound quality.

This conclusion is in agreement with other recent publications on nonlinear distortion in horn/waveguide compression driver subsystems that conclude that virtually all of the distortion in these subsystems is the result of the waveguide itself.[2] The data shown here is substantial support for this position-certainly from a subjective perspective.

One significant result from this study is that there is no reason to consider any aspect of compression driver design from a nonlinear distortion perspective. In fact, one could argue that these results indicate that distortion could be substantially increased, in order to save money or tradeoff other aspects of the design for distortion (like sensitivity), without having a negative impact on the sound quality of the device. Without access to the results presented in this report, the audio community would certainly have taken exception to this conclusion, and would likely object most strenuously.

6 Acknowledgement

7 Bibliography

1. R. Magalotti, C. Zuccatti, P. Pasini, "Building a Plane-Wave Tube: Experimental and Theoretical Aspects", JAES vol. 47, n. 7/8 (Jul/Aug 1999), pp. 596-601

2. A. Voishvillo, "Nonlinearity in Compression Drivers - Where the Distortion Comes From?" 113th AES Convention, 2002, Los Angeles, preprint 5641.

8 Appendix

This experiment uses a *reference stimulus* (the original unmodified musical passage), which is compared to various *test stimuli*. The subject is asked to evaluate audible differences between the various test stimuli and the reference stimulus. This experiment design has its advantages and disadvantages. One advantage is that the task is easier to perform when a reference is continuously available to the subject for comparison against the test stimulus – the judgmental basis is clear and stable. The within-subject variability is thereby minimized, leading to more stability in the results. The negative aspect of including a *reference stimulus* in a design is that the question being asked is to identify "the perceptibility of distortion" and not a "quality level". The reference source design is used to determine which driver is most easily detected as distorted (or different), but it cannot determine which driver sounds best (a quantitative judgment of quality level), hence the use of a reference stimulus is preferred.

If the assumption is that any distortion of the musical signal is undesirable, then the question of detectability provides the complete answer. However, when distortions do exist, and are accepted as inevitable, then it is quite possible that subjects might prefer one driver's type of distortion over another's, which is quite independent of its detectability. This is a subtle but important point, and one that must be dealt with in any subjective testing where quality is a variable.

If, on the other hand, evaluating the preferred sound quality had been the goal (regardless of which source has the most audible distortion and not the case here), then a different test protocol would have been preferred. Among the other possible protocols available is *paired comparison* testing, which is good at evaluating perceived quality level differences in components. In paired comparison tests, different test stimuli are presented to the subject in pairs and the subject is asked to pick the one that they prefer. The end result of this test is a ranking of preference along with the statistical significance of this ranking. The downside of this test protocol is that it difficult to achieve reliable results. Within-subjects variation is often an issue in rankings (intentionally or not-perhaps the result of short term auditory memory) and as a

result, the paired comparison test design typically requires many more subjects and/or trials for a statistically significant sample size. It also has the failing that an inaudible factor or confounding variables simply appear as statistical uncertainty (noise) in the results without any indication as to the root cause of this uncertainty.

9 Figures
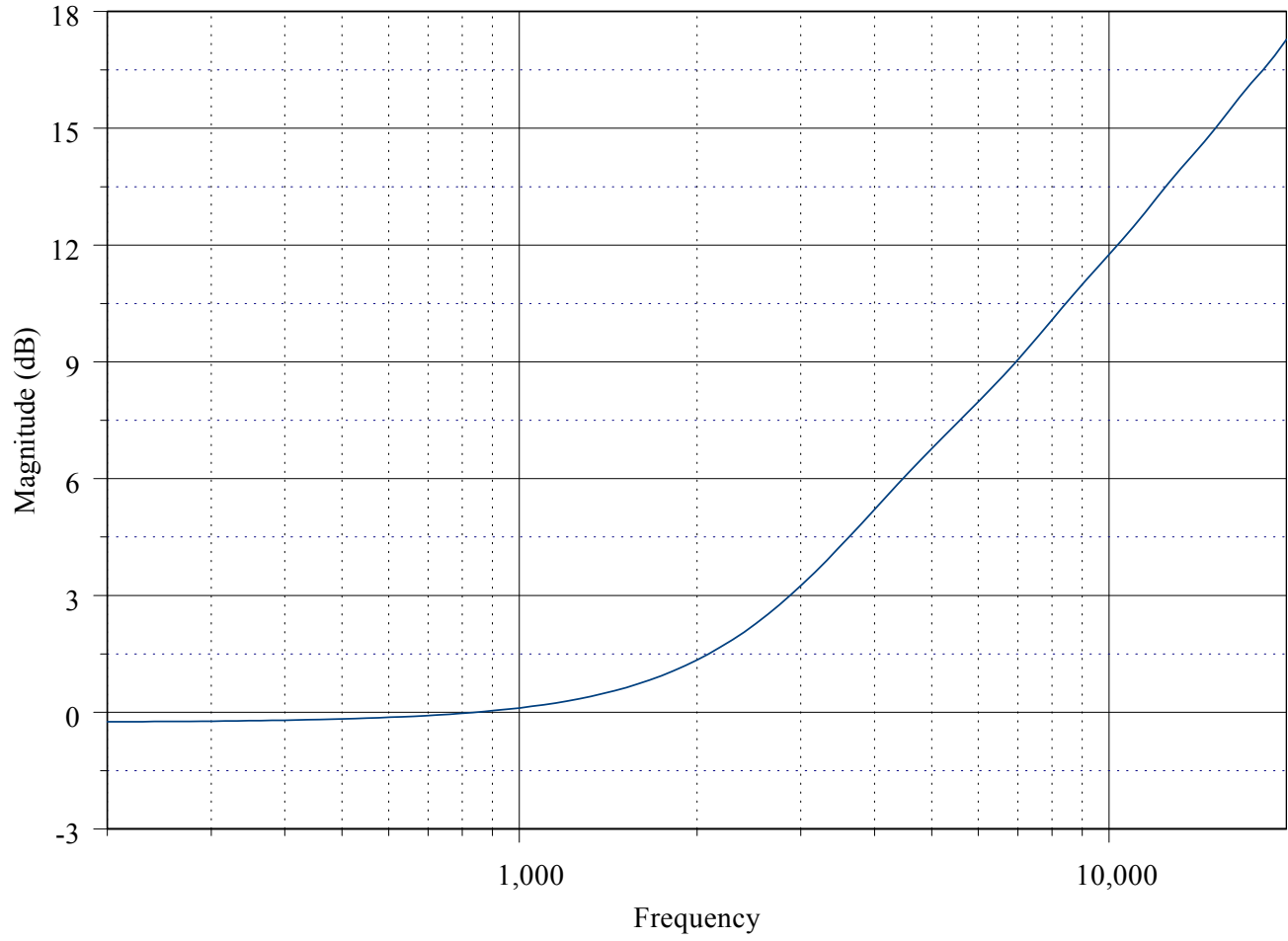

Figure 1 THD for a typical driver
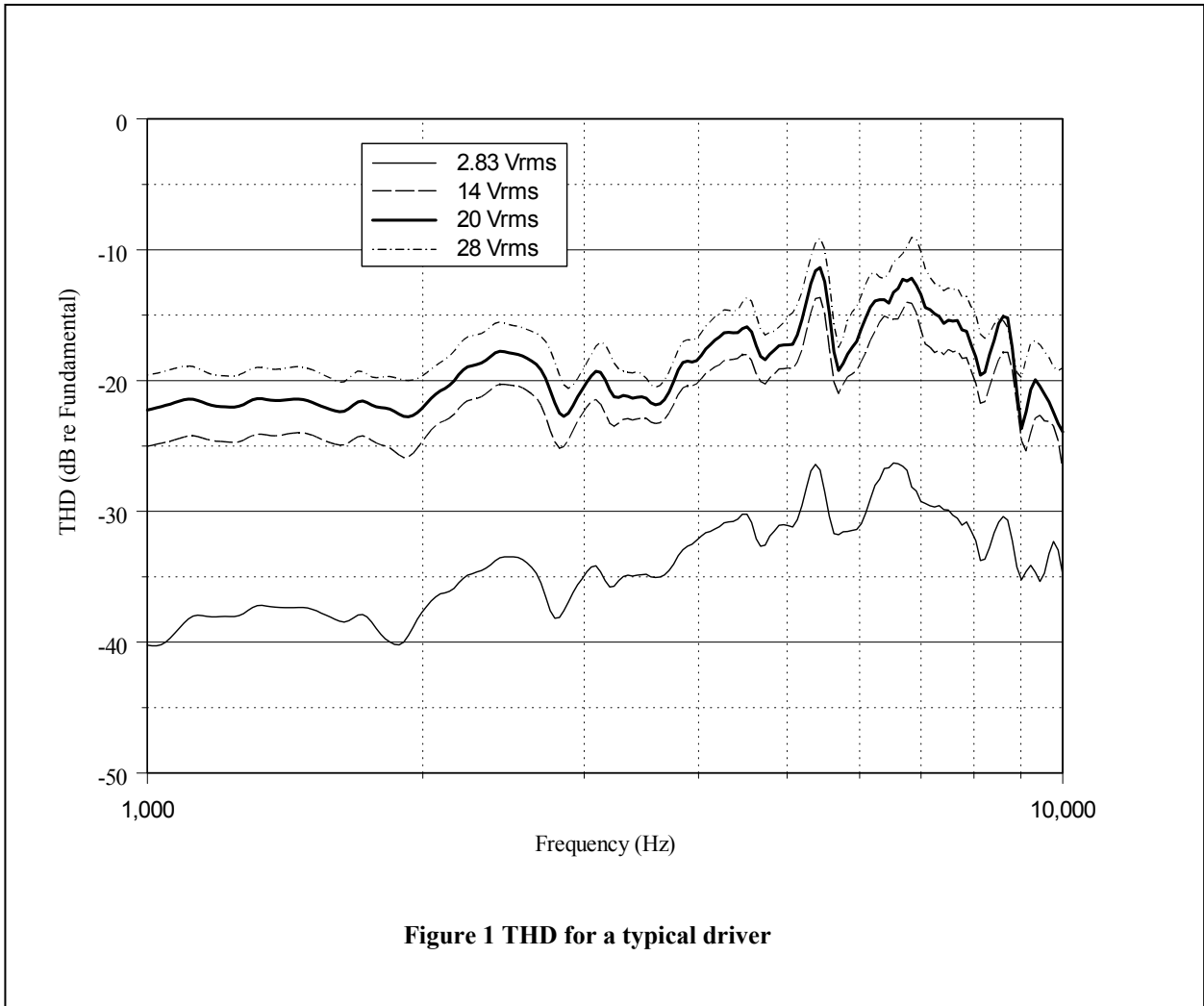
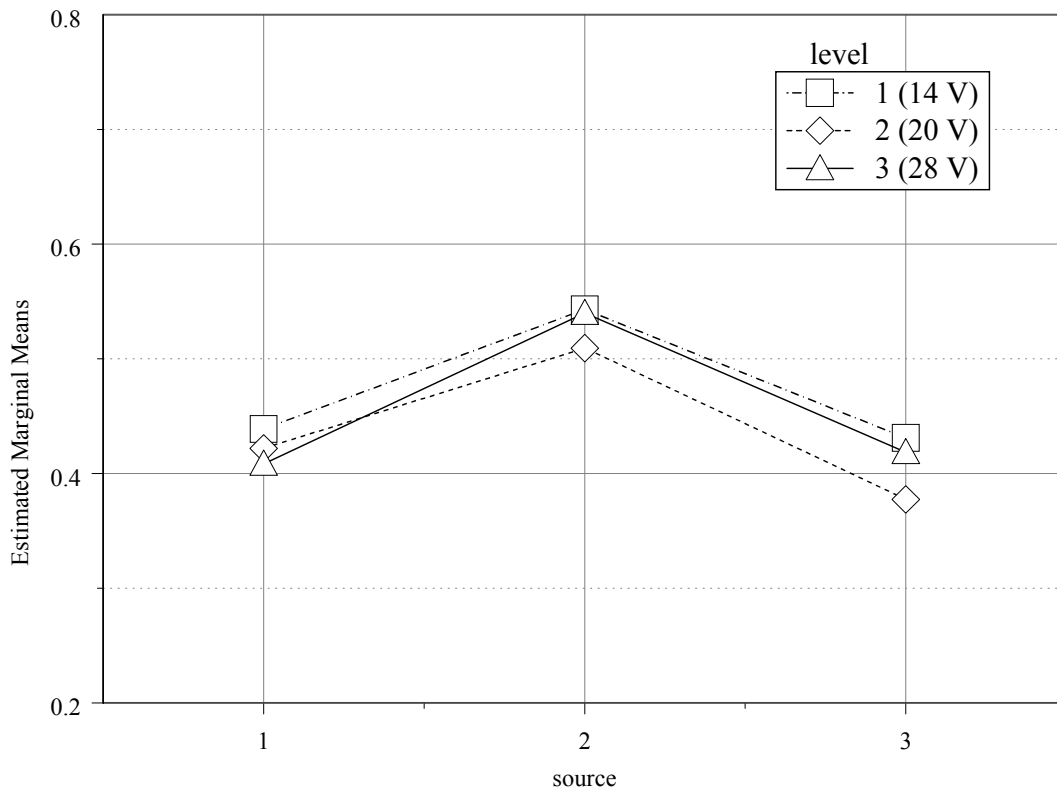**Figure 2 - Frequency response of mass roll-off correction**

Figure 3 –The mean rating across all subjects for Source and Level.