Subjective Evaluation of Reproduced Sound in Automotive Spaces

Roger Shively

Harman-Motive, Inc., Martinsville, Indiana, USA

The automotive space is a harsh and challenging listening environment. This paper describes the listening environment and listening test methods that are used to evaluate the sound quality of an audio system in that environment, and the system design issues involved with each.

INTRODUCTION

In order to understand the motivation for some vehicle listening test methods, it is necessary to understand the environment in which the listening will be performed and the requirements it imposes on the methods as well as the listeners.

VEHICLE LISTENING ENVIRONMENT

As also described in House [1], the automobile is not an ideal listening environment. The one redeeming fact is that the seat position for each listener is known, otherwise the automobile is a small noisy environment with several negative influences on the spectral, spatial, and temporal attributes of a reproduced sound field. A short summary of those negative influences follows.

Spectral.

Many factors can contribute to a vehicle's real and perceived frequency response: The interior volume of a vehicle, the size and shape of interior boundary surfaces, absorption characteristics of all the boundary surfaces, and the location of speakers in the vehicle relative to the listener and relative to nearby boundaries.

For instance, an average mid-sized automobile has a volume of approximately 3.5 m^3. In that volume there will be a large uniform coupling of acoustic modes between 80 and 300Hz. There will typically be some major resonances due to transverse modes between 120 and 150 Hz. The amplitude of these modal resonances can be as high as 12 dB SPL. Above 300 Hz, up to approximately 1kHz, the large coupling of modes is reduced and modal resonances with Q values of less than 3 are numerous. In this region there is a great deal of spectral coloration due to the broader spacing of the modes, and variations in amplitude as high as 8 dB SPL can be observed. The volumes of the automobile's trunk and doors where speakers are mounted can contribute resonances as well between 150 to 500 Hz, with amplitudes anywhere from 4 to 8 dB SPL.

Mechanical resonances can occur due to the vibration of the roof (which is the case for vans and minivans) or the trunk lid due to road motion or engine vibration. These surfaces can radiate acoustic energy at Q values higher than 5 between 300 to 1k Hz with amplitudes from 4 to 6 dB SPL.

Vehicle interiors consist of glass, plastic, carpet, cloth or leather on surfaces that are at various compound angles which respect to each other, the listener and the loudspeaker. Reflections off of these many and varied boundaries result in complex interference and diffraction effects which cause frequency response aberrations and sound coloration at frequencies above 300 to 500 Hz. Interference occurs between

direct and reflected waves. Diffraction will occur through speaker mounting holes and grille coverings, and from mounting locations that are drastically off-axis from the listener.

The vehicle interior also has inherently high ambient noise conditions when the vehicle is on the road. Road, wind, and engine noise can reduce an audio system's dynamic range and mask low frequencies. For the typical mid-sized automobile, the noise is dominant below 500Hz. The average noise SPL is greater than 80dB for frequencies below 100 Hz and velocities above 35 mph.

Some of these problems can be controlled and overcome. Noise masking can be combated with dynamic loudness curves that are tuned to the specific vehicle. Some mechanical resonances can be reduced by reducing the acoustic resonances. Some cannot. The vehicle's structural integrity is the main source of such problems. Some frequency response aberrations can be successfully equalized by an audio system designer. The quality of that equalization work depends on the skill of the designer and how severe the cause of the aberration(s) might be. In general non-minimum phase aberrations cannot be successfully equalized. And, added equalization peaks or dips of Q values higher than 5 can cause phase shifts that can become as audible as other aberrations. The acoustic radiation from a mechanical vibration may be reduced by equalization, but the audio sound may still have a poor damping characteristic, which causes it to sound unnatural.

Spatial and Temporal.

Because of the small dimensions of the vehicle interior, the reverberation time is virtually zero. In this size environment a reverberant field cannot form. The sound field in a vehicle consists entirely of direct energy and early reflections, which are quickly absorbed or dissipated. For the typical vehicle, the decay time for a 60dB signal reduction is 30 to 50ms. This provides a very dead space for sound reproduction. The large amount of early reflections produce a type of diffuse sound field where 90% of the energy occurs within the first 10ms of the direct wave arrival.

[Graph 1(a),(b).] The listener's brain integrates that information and uses it to formulate a perception of the sound field. Effectively the listener is experiencing a diffuse and frontally incident sound field. Looking at impulse response curves or energy time curves [Graph 2.], it can be seen that in this diffuse and frontally incident sound field, those first 10ms of soundwave incidence are somewhat busy. In the driver's position (with the window on the left side of the driver) the time curves show that the sound from the opposite side speaker to the ear (e.g., right speaker to left ear) arrives approximately 1 to 2ms later than the same side speaker to the ear. The arrival from the opposite side speaker has a greater amplitude than the same side arrival. And whereas the same side arrival decays naturally, the opposite side speaker has a reflection off a nearby surface that arrives approximately 4 to 5 ms later before it begins to decay. The amplitude of that reflection is also nearly as large as the initial opposite side arrival. The listener integrates this subtle information and creates an impression of the width and breadth of the sound stage. This is not an experience common with a typical listening room. There are no nearby boundaries to smear the time signature of the sound field and confuse the lateral staging or clarity. With the advent of practical digital signal processors for automotive amplifiers, subtle time corrections, 3D algorithms, and surround sound processing can be used to overcome some of the spatial limitations of the automotive listening environment. This too is at the mercy of the designer's skill and can cause even worse problems if done incorrectly.

VEHICLE LISTENING TEST METHODS

The effects of this "listening room", which is the interior volume of an automobile, on the loudspeakers in that "room" are rather dramatic and severe. The automotive space produces a sound field that is much more complex than the typically controlled listening environment where speaker systems might be evaluated. To make listening tests valuable -- with repeatable, statistically significant, quantitative results which will be useful for a system designer, a

design team, as well as a marketing team --
very careful training of the listeners involved
in automotive sound evaluations is therefore
required. And then, a simple unbiased
listening test method is required.

Listener Training.

The benefits of trained listeners have been
well established for listening in rooms and at
computer workstations. [4],[5],[6] Recent
studies have illustrated the use of a self-
administered PC-based training program to
improve listeners' ability to reliably identify
and rate different types of spectral peaks
and dips which have been added to a variety
of programs (i.e., resonance detection).
[7],[8] Similar investigations have indicated
that critical listening in automobiles might
require a more detailed training regimen and
additional repeats within a trial [9].

The method of using a self-administered
PC-based training program that requires the
listener to identify spectral peaks and dips in
a set of source material reinforces the
listeners' ability to relate a perceived
spectral aberration to a common frequency
scale. However, experience has shown that
because a subject consistently scores high
on such a resonance detection test, it does
not mean that they are well trained for
listening evaluation purposes. Training
sessions in the use of the preference
ranking and timbre balance scales are also
necessary [9]. The objective is to verify that
the listener is applying the resonance
detection training properly in the context of a
listening evaluation or experiment. The
training focuses on frequency-related
problems since these are the problems most
untrained listeners find difficult to describe.

The PC-based resonance detection software
that was used to begin the training is a
homegrown program called EarTrain. The
interface for this software is illustrated in
Figure 1. During each round of training,
there are 24 trials, in each of which the
listener sees 4 different equalization curves
on the screen. There are 6 buttons there,
labeled A, B, C, D, Flat, and Done. Each of
the first five buttons plays a different sound
file. These sound files are the same source
with different equalizations. The Flat button

plays the source with no equalization. The
general idea is for the listener to match up
the equalization curves with the appropriate
sound files, using the Flat curve as a
reference. The first level of the aberrations
was either a 6dB peak or 6dB dip (2.5
octave). After the subjects achieved the
target 95% correct for 6dB aberrations, the
aberrations were replaced with 3dB (2.5
octave) peaks and dips to further refine the
listeners' resonance detection ability.

For the next step of listener training another
homegrown software program called
PrefTest is used. The user interface for this
self-administered PC program is illustrated
in Figure2. This is the same software that is
used in collecting preference information
from our listeners in listening experiments
and automotive sound comparisons. In the
role of a training tool, it is (1) used to
determine agreement among the listeners
for overall preference, and (2) used to
evaluate the timbre balance results to
determine if the listeners are correctly
associating timbre changes with the
appropriate frequency ranges. The same
sound sources that were used in the
resonance detection are used in a
preference testing software. Four buttons
are displayed to the user. These buttons
play the same sources but with a different
equalization. One of the 4 buttons randomly
has a flat response assigned to it for each
trial. This is used as a blind reference.
There are again 24 trials total. For each
equalization, the listener is forced to give a
separate rating number for overall
preference and for the timbre balance of
treble, midrange, and bass. The scale for
the preference is 0 to 10, $1/10^{th}$ pt
increments. The listeners are instructed that
0.5 point is a slight preference, 1.0 is a
moderate preference, and 2.0 is a strong
preference. The scale for the timbre
balances is +5 to –5, 1 pt increments, 0
being neutral.

The listeners repeat the PrefTest training
until they demonstrate a consistent
behavior. If they seem to be having trouble
with any of the concepts associated with
PrefTest the administrator of the tests will
attempt to help them understand better. In
most cases proper use of the scales and
ranking occurs after 2 rounds each of 6dB

and 3dB aberrations. In some cases, consistent but less than ideal use of the scales and ranking occurs no matter how many rounds of training occurred. In either case, the quality of the listener is obtainable in quantitative terms. A group average variance of 0.5 point on a 10 point scale acceptance criterion in the ANOVA of PrefTest results. The individual variance performance is monitored with respect to how they relate to the group and to the blind Flat reference.

Listening Methods.

In-situ sighted listening tests are commonly used within the automotive industry to evaluate the sound quality of automotive sound systems. However, studies on consumer loudspeakers indicate that sighted judgements of sound quality are strongly biased by non-auditory related biases that include size, price, and brand name [10]. It would be logical to suspect that most listening tests done in automobiles may be influenced by these biases as well. Daily work in automotive sound design and evaluations show this to be true, and past experimental results have shown the same indication [9].

Traditionally, blind listening tests in automobiles have not been done for technical and logistical reasons. In-situ blind comparisons between automobiles driven under road conditions are out of the question for obvious reasons. In-situ blind tests with the automobiles stationary are possible but rapid A/B blind comparisons (i.e. paired comparisons) are difficult without the listener having some knowledge of the devices and variables under test. As such, the test can no longer be considered blind. Furthermore, for blind or sighted in-situ tests randomization of variables known to influence judgment of sound, like program material and the automobiles themselves, is also impractical. Yet, without rapid A/B comparisons the discrimination and reliability of subject responses can be negatively impacted due to our limited acoustic memory.

With that in mind, to test just the different methodologies, a method for performing

double-blind in-situ listening tests has been developed and is used as a benchmark for single stimulus comparison of listening methodologies. In this method a strong effort has been made to remove the non-acoustic feedback in the automobile interior. The interior is scented with a very strong scent disk to mask any new-car, old-car, diesel truck, smell that might influence the listener's opinion. The automobile seat is covered with a mat that removes any seat qualities from the listener's opinion. The listener is asked not to touch any part of the interior, but there is always a chance that that request will be violated during the time when a blinded listener is first situated inside the automobile. Therefore, the center console where a stick shift might be is covered with thick sheet of non-reflective material. The foot pedals are covered to remove evidence of a manual or automatic drive. And, the steering wheel is covered to change its feel. The listener is blinded in another room away from the vehicle. A complete blindfold is used to remove any presence of light and visual feedback. Care is taken not to cover the ears. The listener is then given a pair of headphones to wear. Pink noise is played over the headphones to remove any acoustic feedback while the listener is led to the vehicle, helped into the driver's seat, and the door closed. There is an intercom communication fed into the headphones. (Later in the test, the intercom is made available to the listener through a small portable speaker.) After the listener is in the automobile, the pink noise is interrupted and the listener is told to remove the headphones.

Trained listeners with verified good audiometric performance are used as always for the lowest possible variability and highest reliability. The listeners use an interval scale to rate the sound quality of the system in terms of overall preference, timbre balance, spatial fidelity, and absence of distortion. The listeners use the PrefTest software, as illustrated before. In the blind in-situ tests, the administrator runs the software outside of the automobile and communicates with the listener through the intercom. The radio controls for the vehicle are also made remote to the administrator, who maintains a constant system volume and changes source material as instructed

by the PrefTest software, which randomizes the playback. In listening test comparisons where it is not possible to remove or remote the radio controls outside of the vehicle, the administrator occupies the passenger seat and communicates directly with the listener. As a last test in the blind evaluation, the administrator varies the playback volume, and the listener is asked to rate the overall dynamic response of the system from low to high volume levels.

This method is a very reliable single-stimulus benchmark for static systems, and it is being used on a regular basis to perform competitive analyses of stationary automotive sound systems. The method is very consistent, repeatable with meaningful results, and is capable of discerning some subtle differences in a large set of very closely matched sound systems in vehicles considered to be in the same class. Yet a vast improvement in that ability to discern subtle differences could be made if rapid A/B comparisons were possible. And, too, making on-road evaluations is a must for a complete assessment of any automotive sound system performance.

Another method, which does allow rapid A/B/C/D double-blind comparisons to be done in an efficient and cost-effective way, is one that uses a high quality binaural record/playback system.
This method is highly repeatable and allows excellent systematic control of nuisance variables known to influence listening tests.

A binaural dummy head (plus torso and legs) is placed in the exact listening position as a listener would be. Binaural recordings of the source material are made on DAT and edited on a PC as *.wav files. The *.wav files are played back through a digital audio card on the PC, processed through an external D-to-A Converter, and amplified through a headphone amp where the output is maintained at a constant level consistent with the listening level in the vehicle (85dBA). The analog conversion is done outside of the PC to reduce the noise that still exists on most PCs during *.wav file playback. The playback is listened to using Etymotic ER4s earphones, which, with a good seal in the ear canal, provide extremely linear full-band reproduction, at

least 15dB of external sound isolation, and very high repeatability from one session to the next. PrefTest is used to evaluate the *.wav files. The source recordings are randomized and four recordings of a source made in four different cars are presented to the listener, who can rapidly switch between them by clicking on the appropriate GUI button. Recordings can also be made under road conditions and with varying volume settings. The recordings can then be synchronized and presented to the listener who then evaluates the sources under all the different conditions.

This method has proven to produce excellent results when evaluating the spectral aspects of a sound system, whether it is a home system or an automotive system. [3, 4, 11] There is a lack of bone conduction however with the binaural playback method, which effects the absolute agreement of blind in-situ and binaural bass results. They do, however, agree in relative terms. If the bass is ranked higher for one vehicle over the other for a blind in-situ test, it will also be ranked higher for a binaural test, just not on the same scale. [Chart 1(a), (b).] The most difficult point when using the binaural method is that binaural playback of recordings made with a stationary dummy head can prove to be less than accurate in reproducing the spatial aspects of the sound field. Lack of head movement and mismatches between the pinna of the listener and those of the dummy head can cause the results of sound stage and sound image evaluations to be confused when compared to the same results from a blind in-situ test. [Charts 2 & 3.] If recordings using pinna that are similar to the liseners's and recordings related to head movement could be made and synchronized to the listener's head movement, then the playback of the binaural recordings could more accurately reproduce the live experience and become more useful for spatial evaluations.

Another method, which some pilot experiments have shown has very good agreement between it and the blind in-situ method in all aspects of the sound field, is what is called the "placebo" method [12]. The placebo method could be an even more efficient, cost-effective, and controlled

method of sound evaluation than the binaural. In the pilot experiments for the placebo method, as before, qualified listeners used PrefTest to evaluate a single automotive sound system in-situ, blind and sighted. A digital amplifier system was used in the vehicle's sound system. Three equalizations (EQs) were stored in the amplifier's memory. One was the target EQ for the system, and the other two had spectral aberrations added to them. The listener evaluated several sound sources with all three EQs, first blinded and then sighted. The sources and the EQs were randomized, and the listener performed the evaluation five different times for both the blinded and sighted. [3] However, only the data for the target EQ was used as an assessment of the sound quality and analyzed in comparison to the blind in-situ results – with good agreement. Part of the analysis was also to see if the listener was giving different scores to each of the EQs and being consistent in the rankings, and not just ranking them all the same or trying to scale the ranking in order to force an opinion onto the data.

The qualities that the placebo method has over other methods are that it factors out biases and opinions based non-acoustic aspects of the sound. Since the listener doesn't know which EQ is the target EQ, all of the EQs must be given an honest appraisal. And, the listener can do the evaluation in the actual vehicle and while driving on the road. In future experiments the assumption will be made that all systems for evaluation will have a CD or cassette player. The aberrations that were made with the digital amplifier will be made with a re-recording of the source material onto CD or cassette. So, no modifications would need to be made to the original equipment. There will also be more types of aberrations. Not just spectral differences, but small time delays and phase shifts will be added. The intention would be to create a set of sonic aberrations that could be expanded into a larger standardized set of aberrations which could keep growing and changing randomly. With a truly random set of sonic aberrations and enough repetitions to maintain listener reliability, a much better double-blind single stimulus evaluation should be achieved.

The next thing wanting would be the ability to do a rapid paired comparison.

REFERENCES

[1] W.N. House, "Aspects of the Vehicle Listening Environment", presented at the 87th Conv. of the AES, New York, (1989 October), Preprint 2873.

[2] R.E. Shively, W.N. House, "Perceived Boundary Effects in an Automotive Vehicle Interior", Presented at the 100th Conv. of the AES, Copenhagen, (1996 May), Preprint 4245.

[3] R.E. Shively, W.N. House, "Listener Training and Repeatability for Automobiles", Presented at the 104th Conv. of the AES, Amsterdam, (1998 May), Preprint 4660.

[4] Floyd E. Toole, Sean E. Olive, "Listening Test Methods for Computer Workstation Audio Systems", presented at the 99th Convention of the Audio Engineering Society, New York (1995 Oct.)

[5] Soren Bech, "Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment," *J. Audio Eng. Soc.*, Vol 40, (1992 July/August) pp. 590-610.

[6] F.E. Toole, "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance", *J. Audio Eng. Soc.*, vol. 33, pp. 2-32 (1985 Jan./Feb.)

[7] Sean Olive, "A Method for Training Listeners and Selecting Program Material for Listening Tests," Presented at the 97th Conv. of the AES , San Francisco, (1994 November), Preprint 3893.

[8] Sean Olive, "A Method for Training Listeners: Part II", presented at the 101st Convention of the Audio Engineering Society, Los Angeles (1996 Sept.)

[9] R.E. Shively, W.N. House, S.E. Olive [unpublished report] "Pilot Results for: An Examination of Listening Test Methods for Automobiles" (1997 Jan)

[10] F.E. Toole and S.E. Olive, "Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things", presented at the 97th Convention of the Audio Eng. Soc., San Francisco, (1994 November), Preprint 3894.

[11] F.E. Toole, "Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations", presented at the 91st Convention of the Audio Eng. Soc. (1991 October), Preprint 3179.

[12] R.E. Shively [unpublished report] "Pilot Results for: Blind In-situ Method vs. Placebo Method" (1997 Dec)
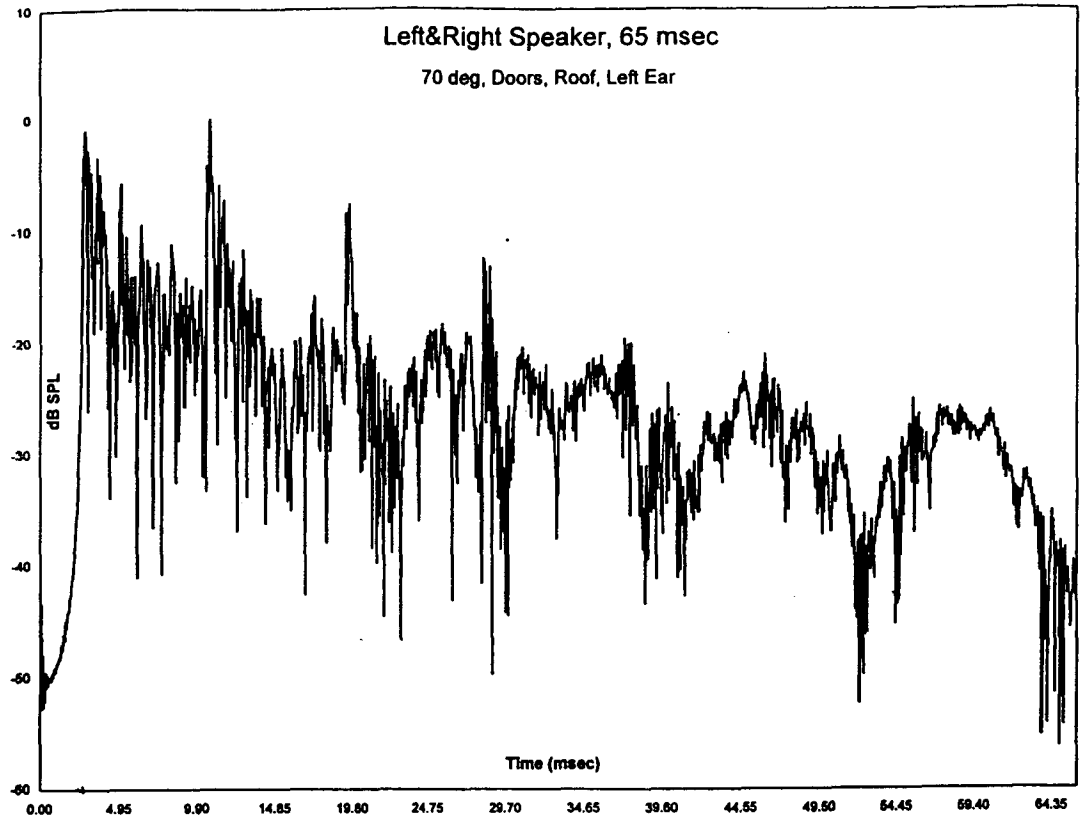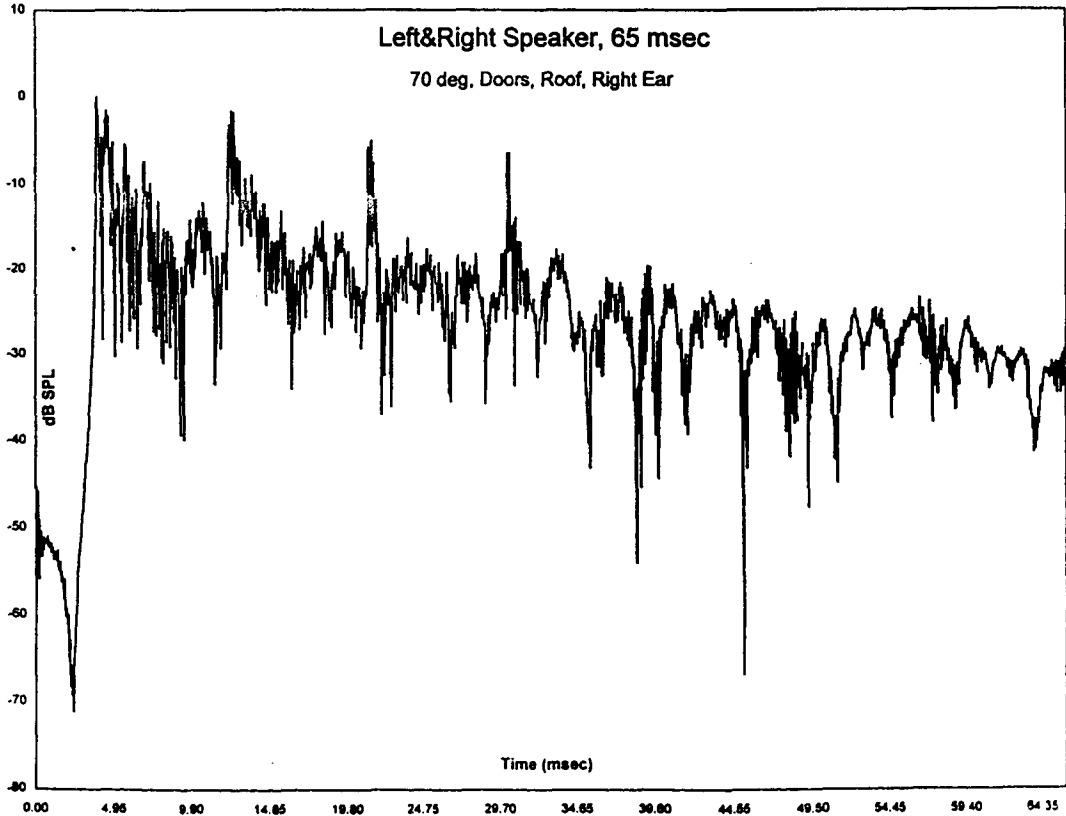
Graph 1. Time Response: (a) Left Ear, (b) Right Ear.
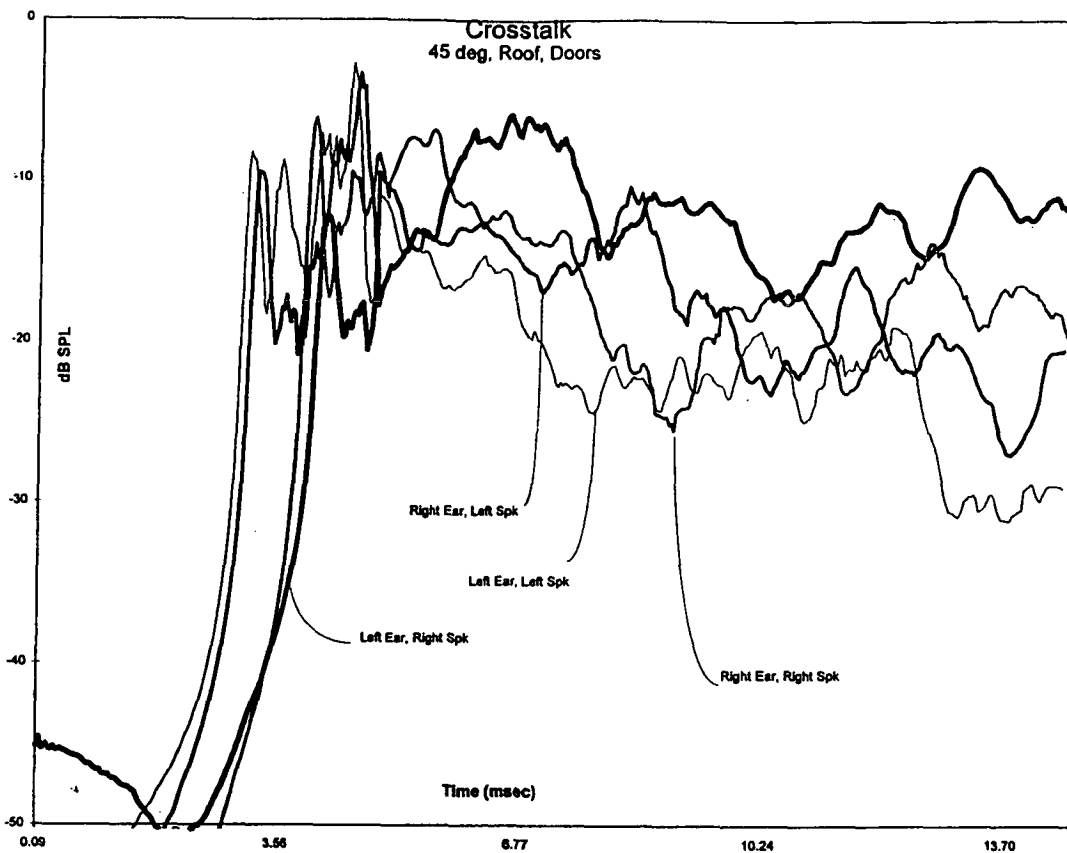
Graph 2. Same-Side Speaker and Opposite-Side Speaker Time Response, Left and Right Ears.
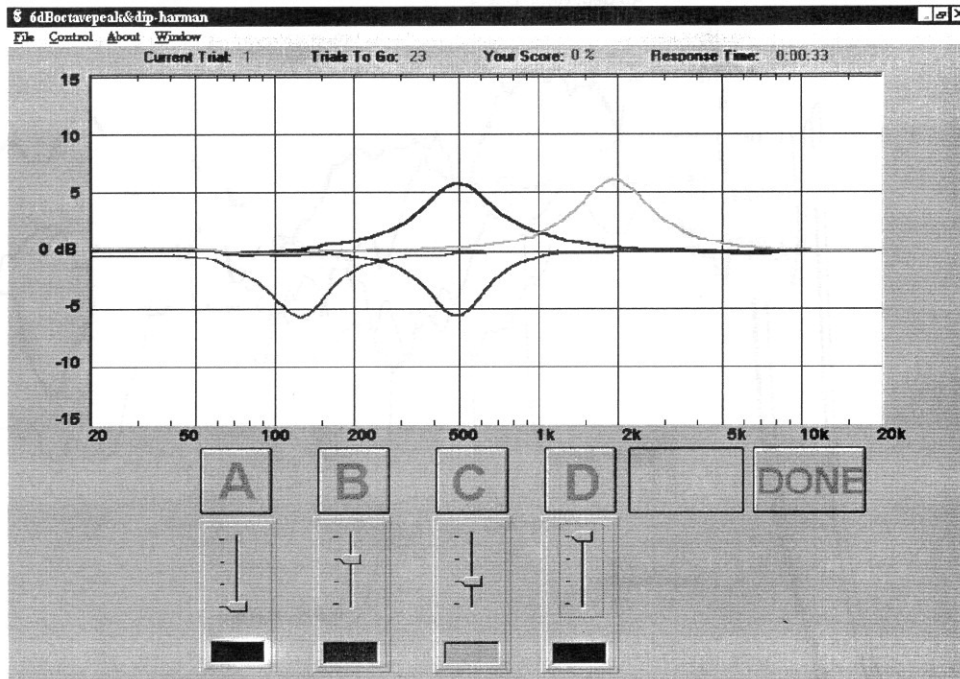
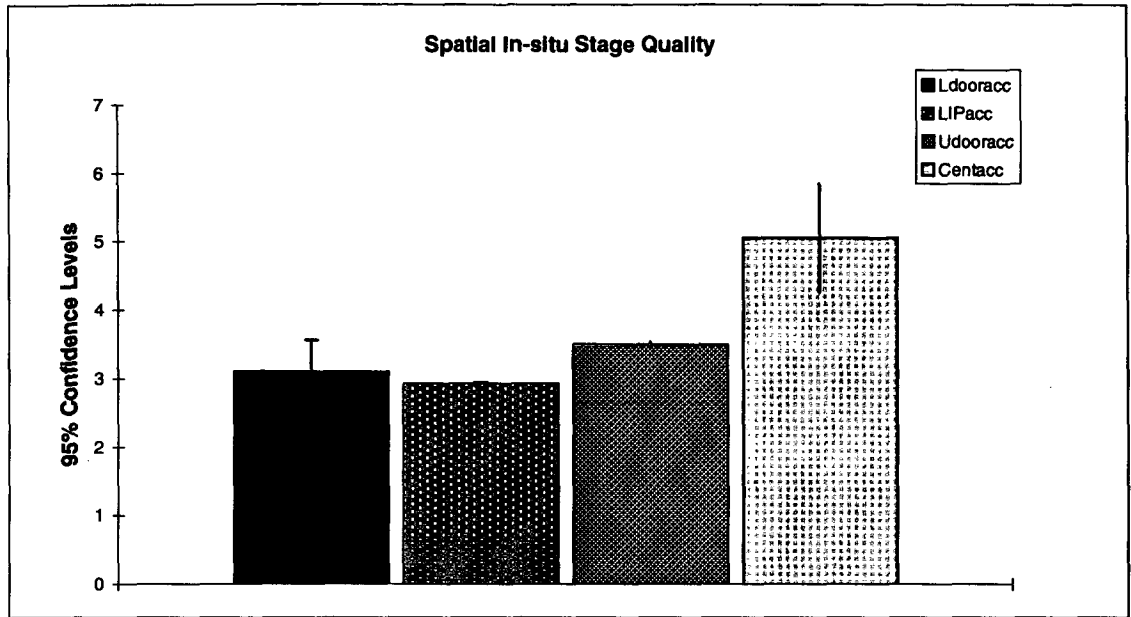Figure 1. EarTrain Software User Interface
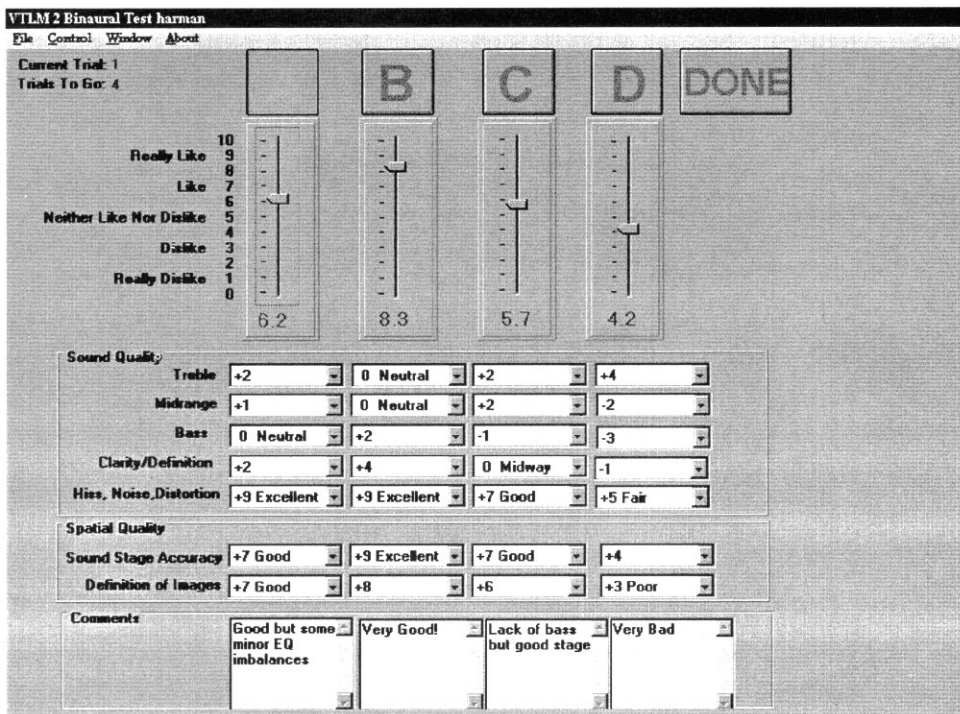
Figure 2. PrefTest Software User Interface.
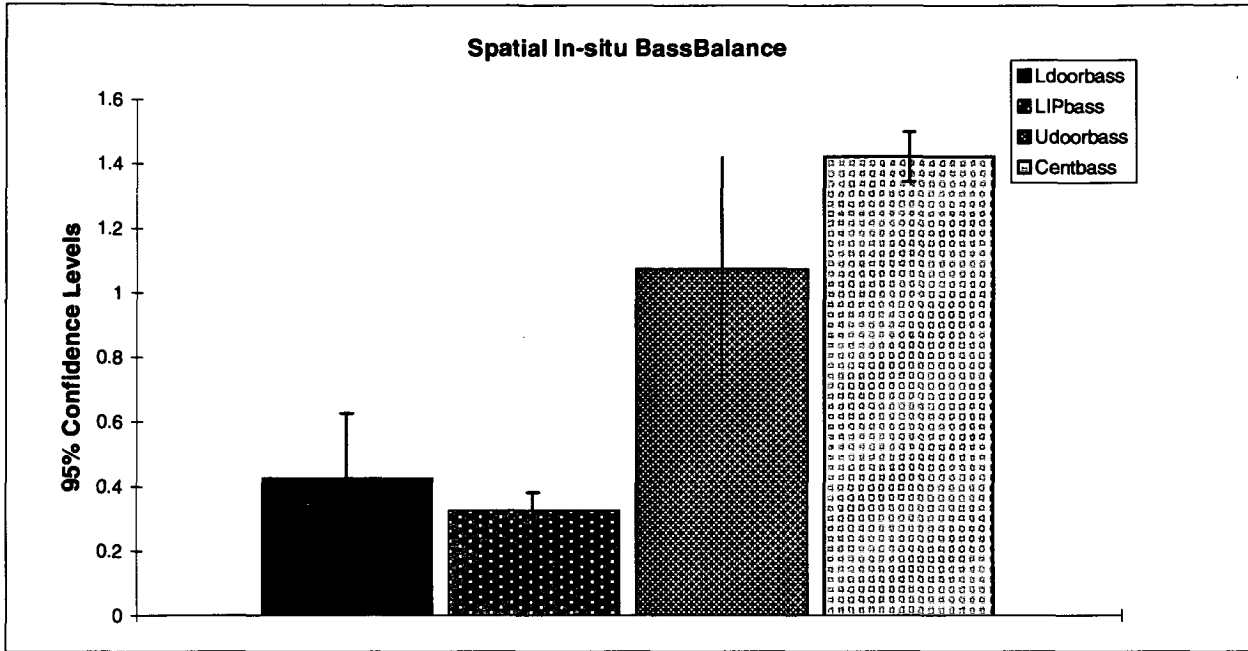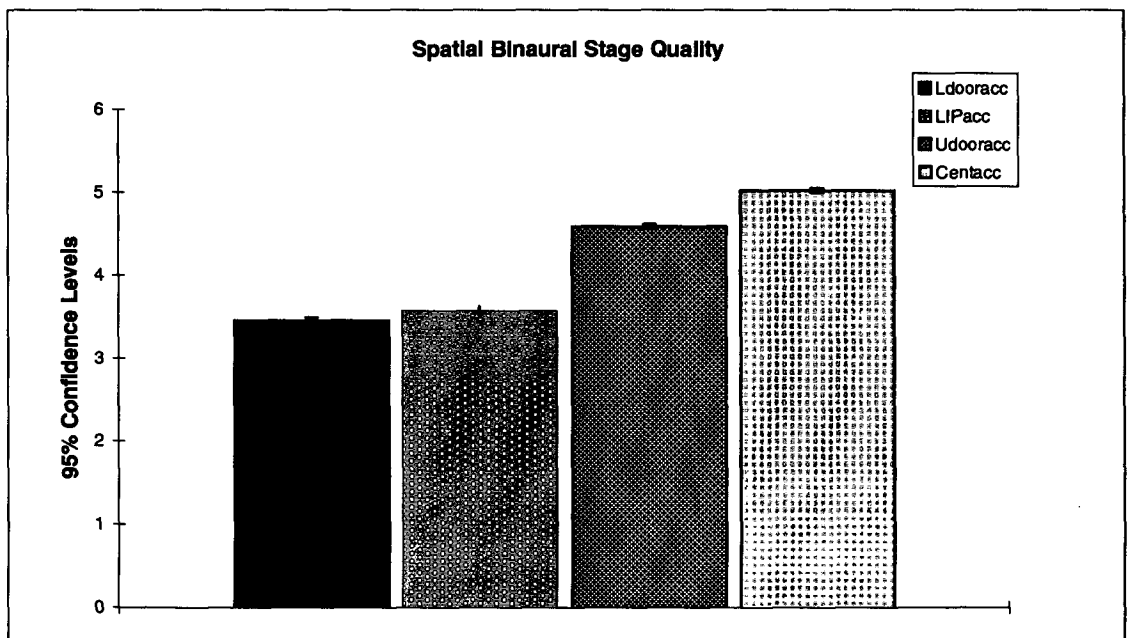
(a)
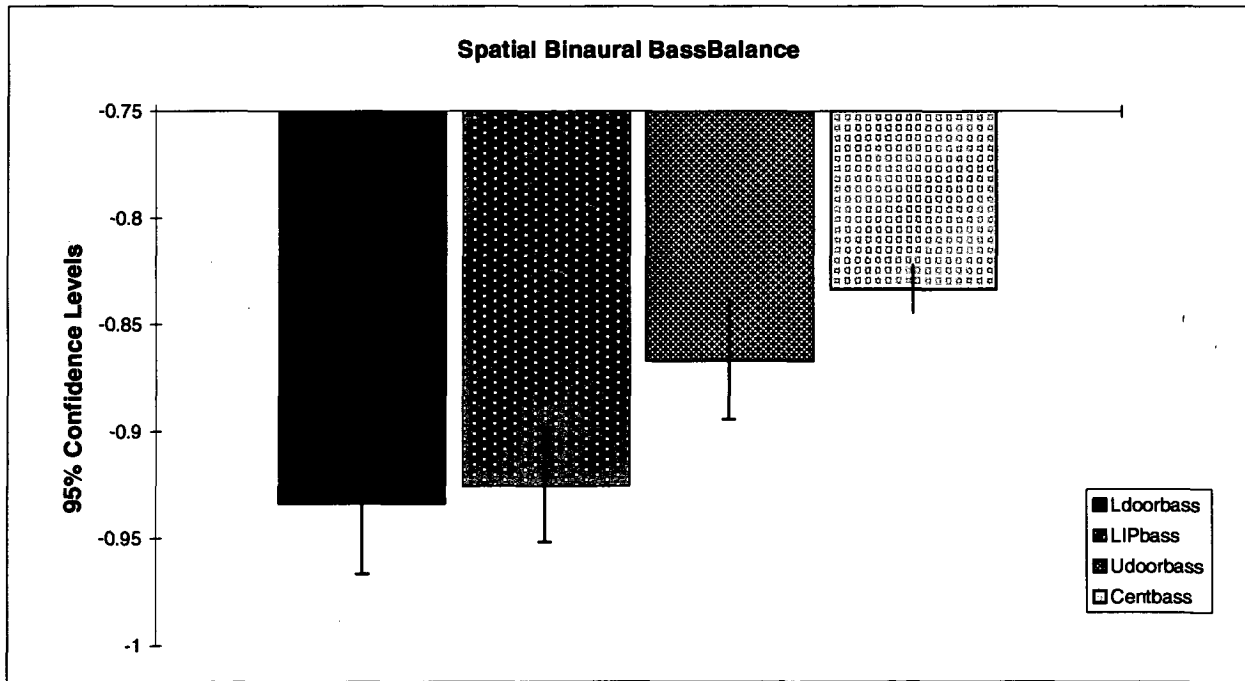
(b)

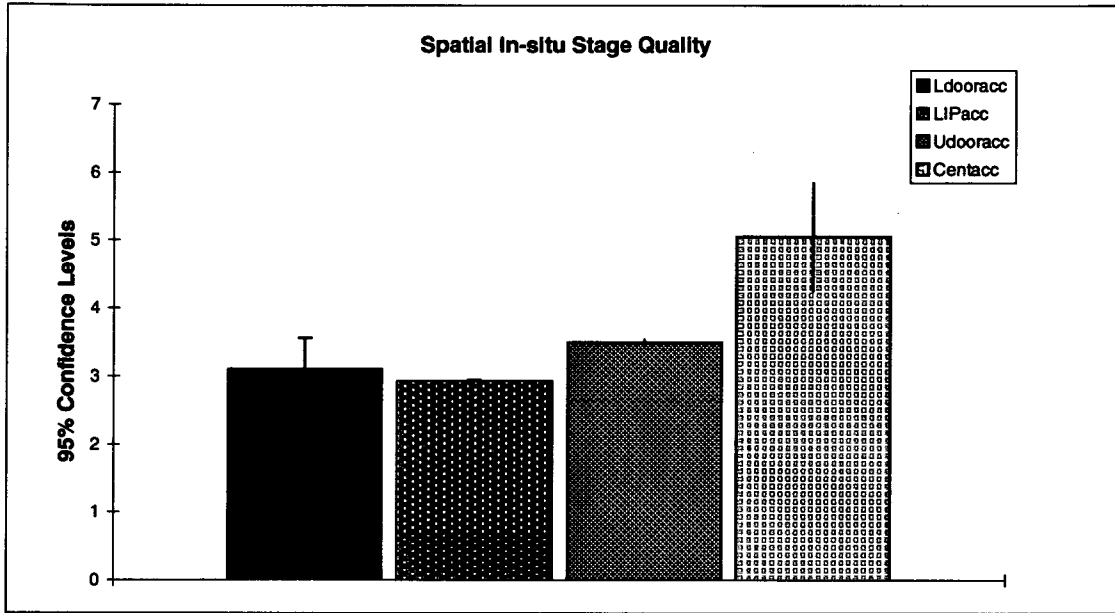Chart 1. Bass Balance: (a) Blind In-situ, (b) Binaural.
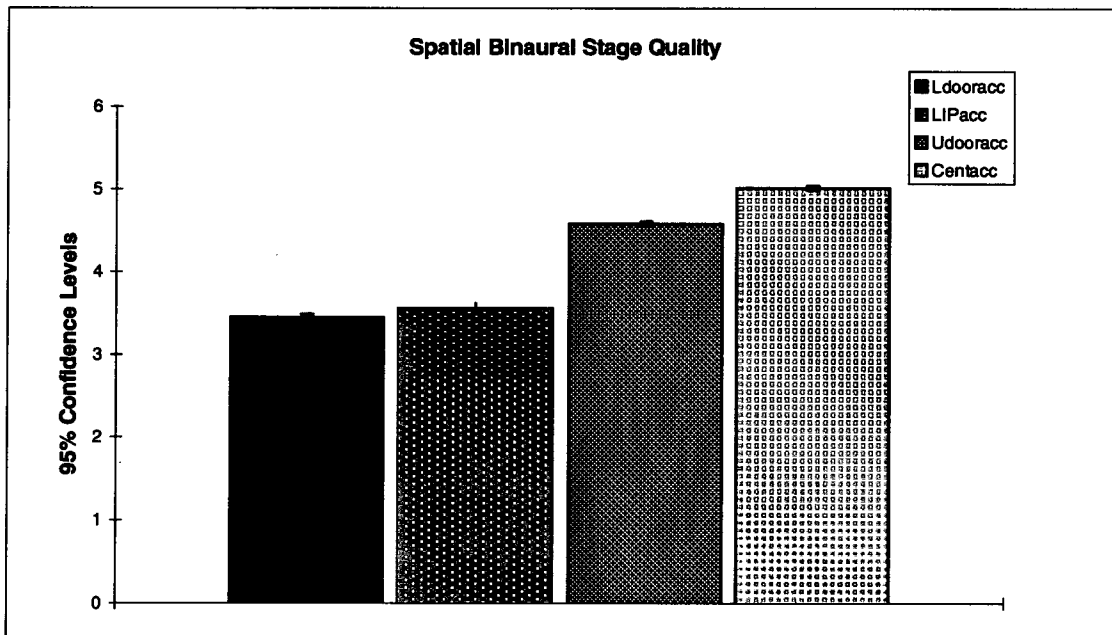
(a)

(b)

Chart 2. Spatial Quality: (a) Blind In-situ, (b) Binaural.

(a)

(b)

Chart 3. Image Quality: (a) Blind In-situ, (b) Binaural.

Page 13