



---

# Audio Engineering Society Convention Paper

Presented at the 110th Convention  
2001 May 12–15 Amsterdam, The Netherlands

*This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

---

## AQT – A New Objective Measurement Of The Acoustical Quality Of Sound Reproduction In Small Compartments

Angelo Farina <sup>(1)</sup>, Gianfranco Cibelli <sup>(2)</sup>, Alberto Bellini <sup>(1)</sup>

<sup>(1)</sup> Università di Parma, Via delle Scienze - 43100 Parma - Italy

<sup>(2)</sup> ASK Automotive Industries, via Fratelli Cervi n. 79, 42100 Reggio Emilia - Italy

### ABSTRACT

The paper describes a new measurement technique of the acoustical quality produced by a sound system. The method is called AQT (Acoustic Quality Test), and produces a graphical representation of the dynamic response of the system to tone bursts at various frequencies. This makes it possible to visualize simultaneously the steady frequency response, the transient response and the signal-to-noise ratio.

The new method revealed particularly useful for describing the performance of a sound system coupled with a small, noisy reproduction space, as it is the case for car audio systems.

### INTRODUCTION

Although many advanced measurement techniques [1,2] have recently been developed for the objective characterization of the acoustical response of sound systems installed in small compartments (i.e. cars), the results of the experiments are usually not easily correlated with the subjective performance. This discrepancy is being addressed with advanced psychoacoustics techniques on the side of the subjective evaluation [3,4,5], but very little effort was done till now in the development of objective metrics which closely correspond to the human perception.

In this paper a new experimental method is described, named AQT (Audio Quality Test), which incorporates in a single measurement and in a single resulting graph both steady-state frequency response and decay/rise time. The original implementation of the method was due to I. Adami and F. Liberatore [6], and is described here altogether with the software measurement tools developed by its inventors and coded by D. Zingoni [7]. The AQT method can be seen as an evolution of the Music Articulation Test Tone (MATT) developed by A.M. Noxon [8].

In the original implementation of the AQT, a test signal is employed made of short sine bursts of increasing frequency, 200 ms long, and with 66 ms of silence between subsequent bursts. The RMS time

history of the system response provides a simultaneous display of the steady-state frequency response (which is the envelope of the maxima) and the dynamic capability (which is given by the modulation depth, or better by the difference between maxima and minima, also called "articulation").

The method was further developed by the authors of this paper, making it possible to post-process with the new technique previously measured impulse responses, and removing the artifact caused by the fact that the depth of the level decrease between subsequent bursts depends not only on the decay rate of the first burst, but also on the rise rate of the second one.

In the modified implementation, the measurement can be done with traditional instruments (such as MLS or continuous sine sweep), and once the impulse response has been measured, the AQT is performed "virtually" by convolution of the test signal with the measured impulse response.

Each burst is convolved separately, avoiding any interaction with the beginning of the subsequent one. The level decay can be assessed simply by getting the SPL value at the fixed time equal to 66ms after the burst end.

Furthermore, also the background noise (due to the engine, tyres and aerodynamic noise) can be included in the "virtual" AQT, avoiding

the need to perform the test while the car is running on the road. The result of the AQT measurement revealed to be very easily understandable, and to correspond well with the subjective perception, making it possible to pack in a single graphical description the three concepts of frequency response, decay time and signal-to-noise ratio. The paper includes detailed description of the two implementations (real time measurement and virtual measurement obtained by convolution), some samples of measurement results obtained with different car sound systems, and a subjective/objective comparison of the results.

**THEORY**

The theoretical background of this work is known since a long time, and thus it will be recalled here only very briefly. When an electroacoustic sound source is employed for reproducing a signal in an enclosed space of limited dimensions, the sound produced by the loudspeaker interacts strictly with the sound field, and it is not easy to separate the effects caused by the loudspeaker from the effects caused by the room. The smaller the room, the more strict is this interaction: thus car sound system are the worst case, and this explains why loudspeakers which sound fine in an anechoic room are bad inside a car, and vice versa. In large rooms, which were studied much more extensively in recent years, it is common to analyze the sound field by means of a room impulse response measurement. Observing it, it is usually possible to isolate a direct wave, a subsequent delayed packet of discrete reflections, then a mixing zone (also called cluster), and finally a smooth, statistically-behaving reverberant tail. None of the above distinctions makes sense in a small room or inside a car: direct sound and reflections mix together in a continuum, decaying burst. Looking at the steady-state frequency response, obtained simply by taking the FFT of the whole impulse response, the behavior is highly uneven at low and medium frequencies, due to strong resonances and evident acoustic modes of the cavity. These are very well separated at low frequency, the modal density increases with frequency, but due to the small dimension of the acoustic space the modal density becomes really large only at high frequency, where the response begins to smooth due to the interaction of adjacent modes.

It is usually accepted that the modal behavior and the “small room” effect are limited to frequencies not exceeding 2 kHz, and thus in this work the low and medium frequency ranges are explicitly addressed. The things are even more complex taking into account that the sounds to be reproduced are usually constituted by music and speech. Both these signals are strongly not-stationary, and their proper perception is severely affected by the transient response of the electro-acoustic and acoustic system. Again, a lot of experience is available for large rooms (i.e. concert halls), where many objective parameters were defined for describing the transient response of these large acoustic spaces. Among these parameters, we could cite the most famous ones, such as the reverberation time (T20 or T30), the Center Time  $t_c$ , the Clarity  $C_{80}$ , the Definition  $D_{50}$  and the Strength G. All these parameters are defined and standardized in the ISO3382/1997 measurement standard, which also gives details on the way of measuring room impulse responses and of deriving these numerical parameters by digital techniques. The authors developed specific software tools for measuring ISO3382 parameters [1]. None of the above parameters revealed to be subjectively relevant in the assessment of sound systems in small spaces. The reason is that the time constant of the acoustic space is usually not much longer than the time constant of the electroacoustic transducers, and thus the two decays are mutually coupled. Furthermore, in the absence of any appreciable delay between the direct sound and the reflections, our hearing system is not capable of separating them, and thus the human perception is governed more by this interaction than by the free-field behavior of the transducers.

What is needed therefore is a measurement method capable of resembling what happens in the human hearing system, and specifically assessing what’s the response to short transients of sound with limited frequency content and repeated at a sustained rate (this can be seen as a simplistic description of a music piece). This original approach was first exploited by I.Adami and F. Liberatore [6], who developed the whole methodology nowadays known as AQT, and published it in the format of a software for the Windows platform called Sound Analyzer 4.0 [7], implemented with the help of D.Zingoni.

**THE BASIC AQT METHOD**

The basis of the measurement technique developed by Adami & Liberatore is the generation of a special test signal, which have to be reproduced in the space to be assessed. This signal is a repetition of sinusoidal bursts, with a length of 200ms, and intervalled by 66 ms one after the other. The frequency is slightly increased among the previous burst, with a step of 2 Hz starting from 20 Hz up to 300 Hz, then the step is increased to 4 Hz up to 1000 Hz, and it is further increased in exponential steps above 1 kHz and up to 2 kHz. It can be noted how this frequency spacing roughly resembles the BARK scale.

Fig. 1 shows a close-up detail of a couple of these bursts, which are properly faded in and out through linear amplitude transitions.

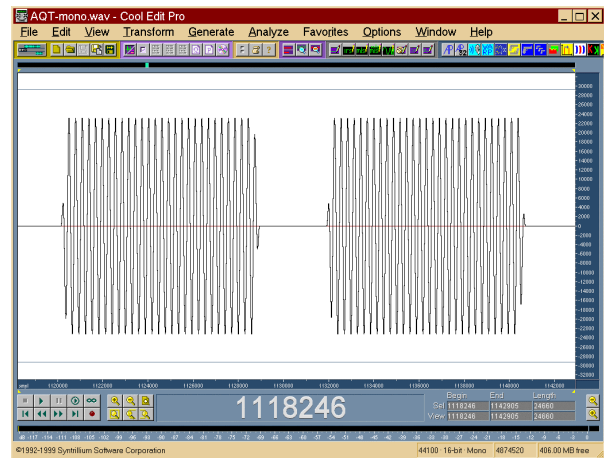


Fig. 1 – sinusoidal bursts

It must be noted, anyway, how these very rapid amplitude transitions at the beginning and end of each burst are prone to produce some frequency-domain artifacts (“leakage”), as clearly demonstrated from the sonograph visible in fig. 2.

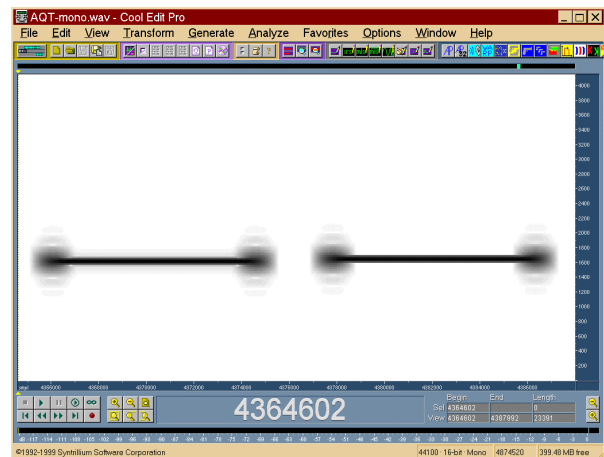


Fig. 2 – frequency leakage at the extremities of each burst

Of consequence, when the whole frequency span is taken into account, it appears that the sequence of burst excites the system not exactly one single frequency at once, but with a sort of frequency-domain filter, having an aperture of approximately 1/6 of octave. Fig. 3 shows the complete spectrogram of the entire test signal.

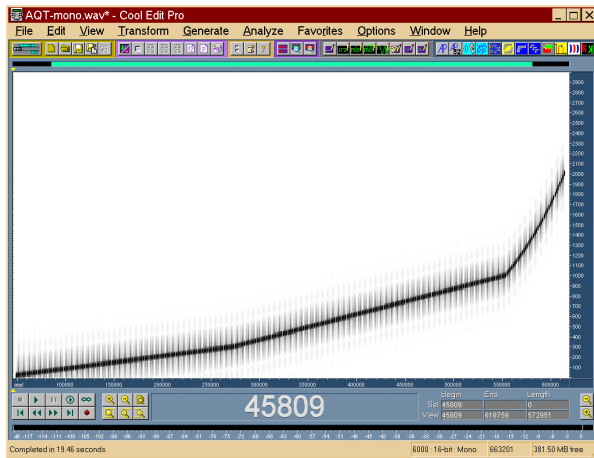


Fig. 3 – sonograph of the entire sequence of bursts

While this special test signal is being played, at the listening point a temporal recording of the instantaneous RMS level is taken. The time constant for the RMS computation is 25 ms, as this is considered to be the time constant of the human ear.

The graphical plot of this level recording is already the AQT measurement result: as it is seen, the method is in principle very simple, and does not requires any advanced numerical signal processing of the recorded sound.

Acustica Applicata also produced a specific software tool [7], running under Win32, for automating the task of generation of the test signal, measurement and plotting of the RMS level vs. time, and subsequent post processing for deriving a quantitative metric of the dynamic capabilities of the system (the AQT value).

Fig.4 shows the user's interface of the main level recording window of this software, which is called Sound Analyzer [7].

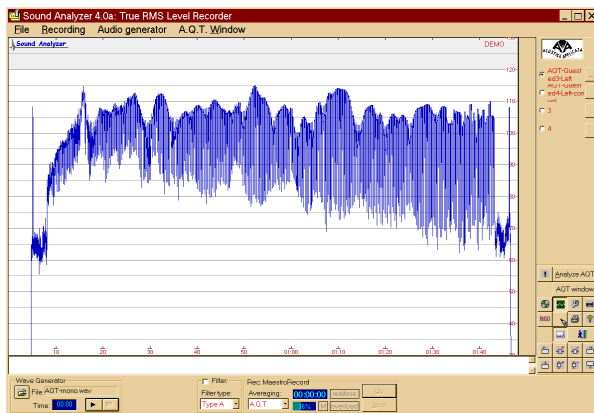


Fig. 4 – Level Recorder window

It can be seen that before and after the sequence of bursts, two sharp spikes appear: these are synchronization signals, which are simply very short bursts at 1 kHz.

Although some information can be seen also from this level-vs-time graph, the software is provided with a post-processing window, where each burst is individually selected and processed, and the results are displayed in a level-vs-frequency plot, as shown in fig. 5.

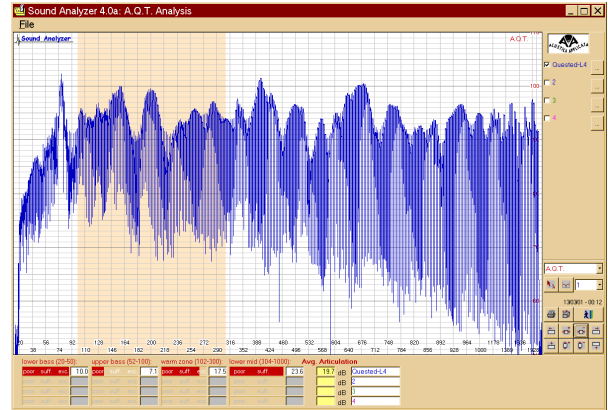


Fig. 5 – the AQT window

A first look at fig. 5 shows that at a very low frequency (74 Hz) the room has a very strong resonance, which produces a peak in the steady-state response, and a very little value of the level variation at the end of the burst.

The frequency axis was divided in four zones, which are called respectively:

- 20 – 50 Hz : lower bass
- 52 – 100 Hz: upper bass
- 102 – 300 Hz: warm zone
- 304 – 1000 Hz: lower mid

the fifth zone, 1004 to 2000 Hz, although not being explicitly addressed, is taken into account for the derivation of the wide-band, single-value AQT.

It is possible to zoom on the frequency axis, revealing the detail of the closely-spaced bursts, as it is shown in fig. 6.

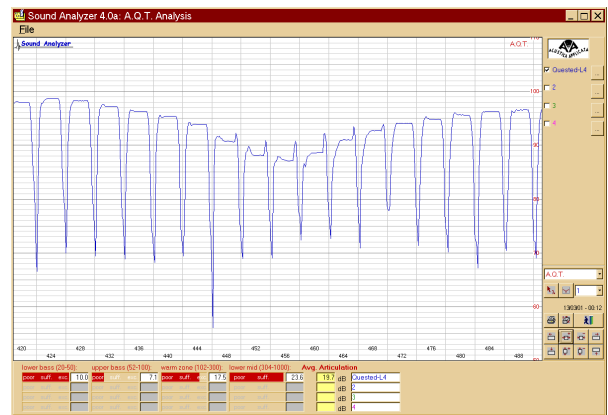


Fig. 6 – zoom on the frequency axis

From fig. 6 a typical behavior can be seen. At the receiver's point, on the left part of the figure, the sound being reflected by the room is substantially in-phase with the sound being directly radiated by the loudspeaker. In fact the sound level first reaches a value of approximately 95 dB, but then it continue rising, up to slightly more than 97 dB. The bursts thus appears rounded and without spikes.

In the center of the figure 6, instead, being the frequency slightly increased, the opposite happens: the sound coming from reflections has opposite phase respect to the direct sound. Although the initial part of the bursts still reaches a sensible level, when the subsequent reverberant sound field do establish, the RMS level is reduced, because these reflections are interfering destructively with the direct sound. The bursts now present two very evident spikes (also called overshoots) at the beginning and at the end, which happens when only the direct sound is present (at the beginning) and only the reflected one is present (at the end).

The numerical evaluation of the AQT is not based on these facts. Instead, the quantity which is taken into account is basically the level decrease which occurs between two subsequent bursts.

After a strong sound, the human hearing mechanism exhibits the well-known effect called masking: this means that, even if the acoustical level falls very quickly, our ear is not capable of detecting small sounds which occur just after a strong one. This phenomenon was first studied by Zwicker [9], and fig. 7 reports his widely employed graph, demonstrating how after a burst of a certain duration the sensitivity is reduced.

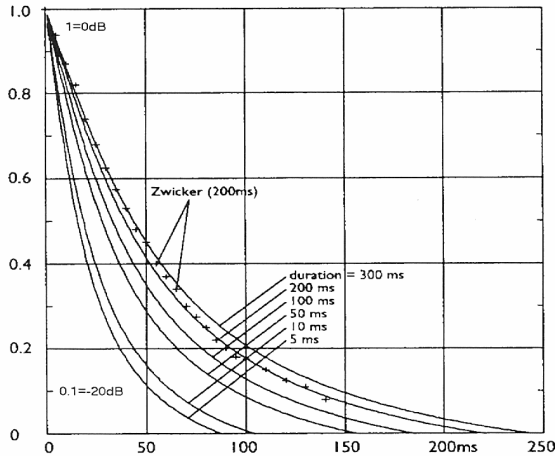


Fig. 7 – time-domain masking

Observing the curves above, we note that after 66ms the masking curve is approximately at -20 dB. This means that, even if the acoustical level falls more than 20 dB between two bursts, we must consider an upper limit of 20 dB of subjectively perceivable amplitude modulation (also called “articulation”). When instead the depth of the amplitude modulation is lower than 20 dB, this means that we are experiencing an evident “sustain” at the end of each note, due to a relevant amount of reverberating energy “attached” just after the end of the burst.

It is important to understand, at this point, that this short-term tail has nothing to do with the (much longer) tail which is typically present in concert halls and other venues with rewarded acoustics: in these large rooms the direct sound is always followed by a certain gap of silence, and only after 15-30 ms do the first reflection occur. This means that in a large space each note (or phoneme) is always completely terminated when the subsequent reflections arrive, and although our brain do not interpret them as echoes, it is indeed capable of retrieving almost unaffected the information carried by the direct sound. This low-term tail, instead, is quite detrimental to the dynamic perception of the sound, it smears the transients and makes the sound dumb.

So the goal is to obtain a modulation depth as large as possible, within the perceivable limit of 20 dB. The evaluation of the modulation depth is done for each single burst, and the results are averaged for the 4 main frequency regions and for the complete frequency range, as shown at the bottom of figs. 5 and 6.

It is also possible to look at a single burst, and see what is the modulation depth at a very specific frequency. Selecting this function, the Sound Analyzer software shows the window reported in fig. 8.

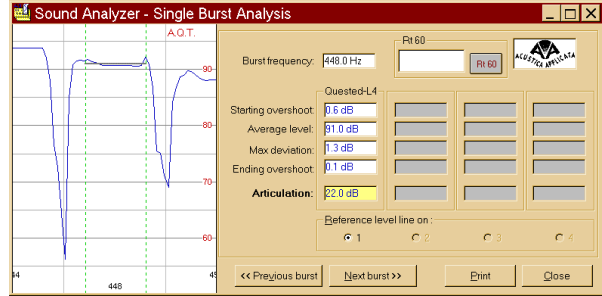


Fig. 8 – single burst analysis

It must be noted, anyway, that the minimum level reached at the end of the burst depends not only from the decay of the burst itself terminated, but also on the attack of the subsequent one. Furthermore, also the background noise could intervene into limiting the modulation depth. In this sense, the AQT analysis seems to exhibit some points very common with the Modulation Transfer Function analysis by Houdgast and Steeneken [10].

**THE VIRTUAL AQT METHOD**

This first extension onto the original AQT method was obtained employing the capability of measuring the system’s impulse response and convolving it with any given signal, obtaining a realistic simulation of the system’s response. This is possible making use of the software tools developed by one of the authors [1].

This makes it possible to substitute the measurement procedure employing the multi-burst test signals, which is quite long and very sensitive to impulsive noise events, with a much faster and reliable impulse response measurement, which can be done both with the MLS method [1,11] and with the sine sweep method [2]. After that the system’s impulse response has been measured, the AQT analysis can be performed on a synthetic Wav file, obtained convolving the AQT test signal with the measured IR. Fig. 9 shows, for example, the measured IR of an high-level loudspeaker (Questaed VS2108) installed in the ASK listening room and measured with the MLS method.

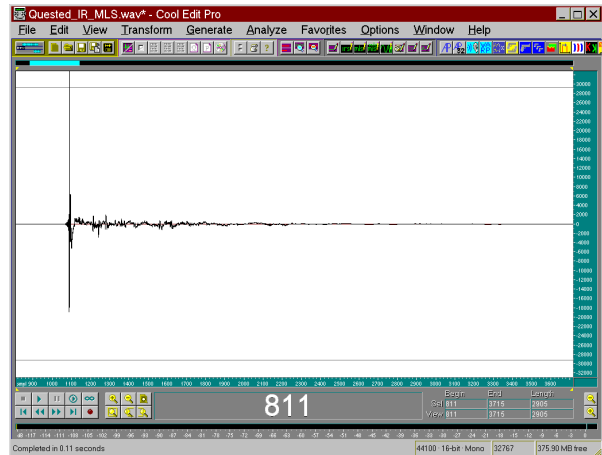


Fig. 9 – measured IR of a Questaed loudspeaker

Fig. 10 shows the comparison between the waveforms obtained by a live recording of the AQT test signal, and its synthetic counterpart, obtained by linear convolution of the test signal with the above IR.

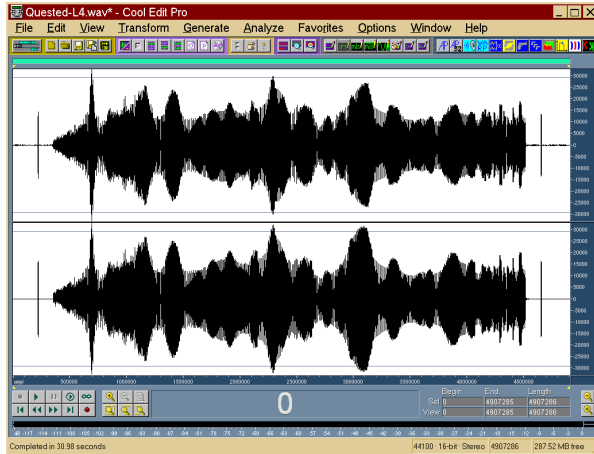


Fig. 10 – live recording (top) and synthetic convolution (bottom)

By careful inspection, and by listening at the sound, it appears that the synthetic signal is somewhat different only during the silence before and after the sequence of bursts, where it is much more clean, missing any background noise. But the transients evoked by the burst are properly reconstructed, even in minor details.

After the creation of the synthetic signal, a “virtual” AQT analysis was performed on it. The results were compared with those obtained by a “live” AQT analysis of the same sound system, and the results of this comparison are shown in fig. 11.

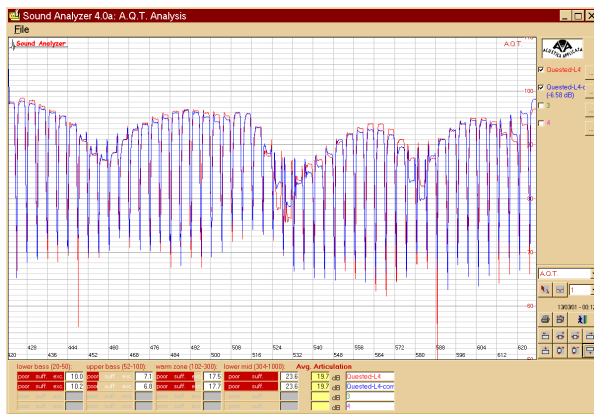


Fig. 11 – comparison between the AQT analysis of the live and synthetic signals

It can be observed how the profiles are close each other, although some very minor deviation do appear: these deviations are of the same magnitude which appears repeating two “live” AQT measurements, and thus are not of any practical significance. The average values of the modulation depth are substantially the same in the 4 frequency regions and in the wide-band analysis.

We can conclude that the convolution approach is the natural complement of the AQT analysis, making it possible to conduct the measurement in much less time, and obtaining more robust results.

**AQT2 – THE NEW METHOD**

Although the synthetic approach already removed some of the most evident problems with the original AQT method, the authors developed a more advanced analysis method, which was called AQT2.

First of all, a new stand-alone application was created. Both the Graphical User Interface and the computational engine were written in Matlab, and the resulting application was compiled for gaining in speed and portability.

Fig. 12 shows the main windows of the new application, displaying the smoothed magnitude of the frequency response of the same loudspeaker already analysed in the previous chapter.

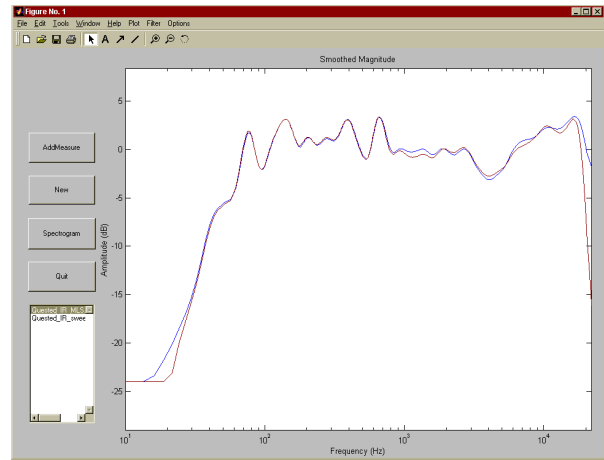


Fig. 12 – smoothed magnitude of the frequency response of a Quested loudspeaker, measured with MLS and Sine Sweep methods

From the analytical point of view, the main difference between the new method and the original one is that each burst is processed separately from the others, avoiding the problem of the interaction of the tail of one burst with the attack of the subsequent one. This was possible generating many separate bursts, each 200 ms long, preceded by 50 ms of silence and followed by 300 ms of silence. Each burst is linearly convolved with the measured IR, which is loaded in the software from a standard Wave file.

The result of the convolution with all the bursts is thus constituted by a sort of sonograph, in which the transient response of the system can be seen at all the frequencies simultaneously, with much more details than what can be seen from the level-vs-time graph. Fig. 13 shows this kind of sonograph.

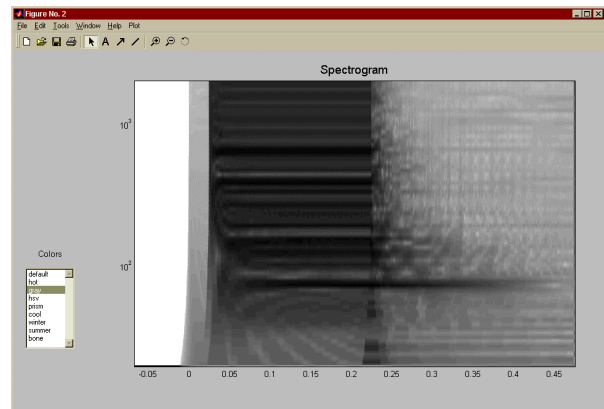


Fig. 13 – sonograph of the transient response

In the above picture, it is easy to see the long reverberant tail which occur at the room’s cavity resonance (74 Hz). Clicking over the sonograph at that frequency, the single-burst analysis window appears, as shown in fig. 14.

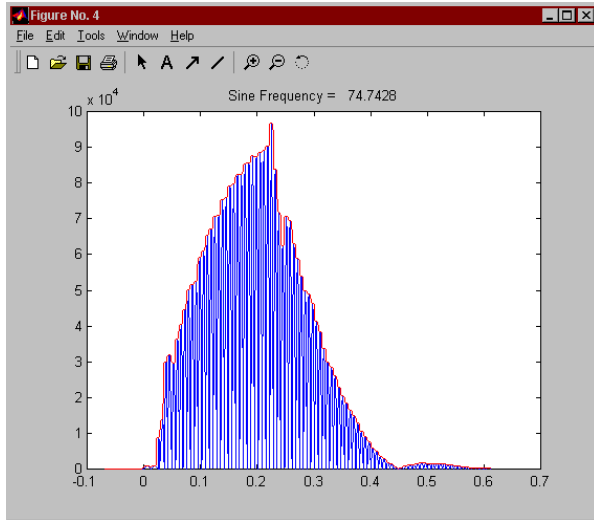


Fig. 14 – single burst at 74 Hz

It can be seen how, at this frequency, the system was very slow in reaching the steady state. After 200ms it was yet increasing, when the burst ends, and a very long decay starts. For comparison, fig. 15 reports a single burst at the frequency of 392 Hz, where instead the system is very quick in following the burst shape.

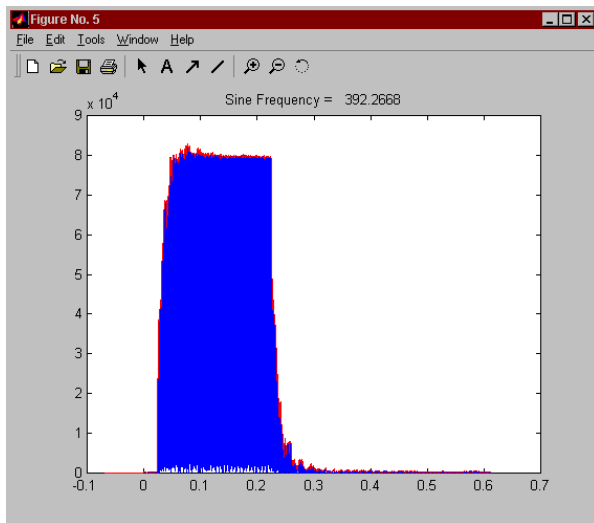


Fig. 15 – single burst at 392 Hz

Finally, it is interesting to see what happens in the regions in which the phase of the reverberant field is opposite to the phase of the direct sound (as already discussed regarding fig. 6). Fig. 16 shows one of these cases, at the frequency of 460 Hz. Also in this case, very evident spikes appear at the beginning and at the end of the burst.

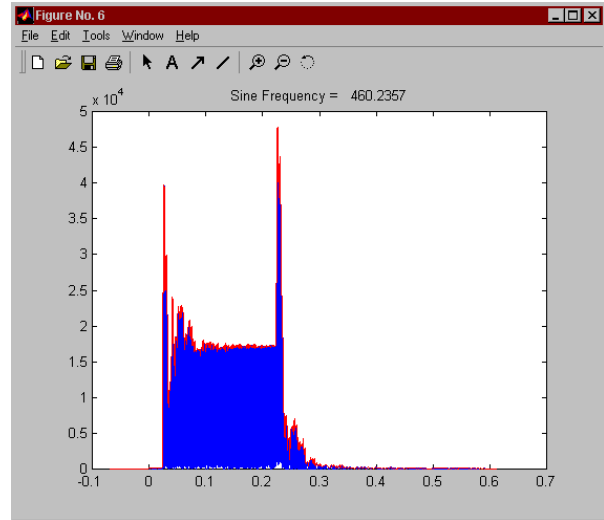


Fig. 16 – single burst at 460 Hz

The observation of these strong overshoot phenomena suggested that it is important not only to measure and evaluate the steady-state frequency response (which corresponds to the level in the central plateau between the two spikes), but also the overshoot response, that is the maximum RMS level reached at any frequency in correspondence of the burst extremes.

For each single burst (and thus for each frequency), a numerical analysis was performed, which yielded the following 4 values:

- Max Overshoot level
- Steady State level
- Decayed level after 33ms from the burst end
- Decayed level after 66ms from the burst end

It was chosen to take into account also the level after 33ms, because this is thought to be more significant for car sound systems, where the high background noise level limits the dynamic range to much less than 20 dB. These 4 level-vs-frequency curves are plotted on the same graph, which this way describes completely the measurement results.. Fig. 17 shows this multi-curve graph for the Qusted loudspeaker installed in the ASK listening room. In this case, the curve at 33ms after the burst end was not displayed, both for not making the graph too complex to read and because in a listening room the 66ms value is more significant.

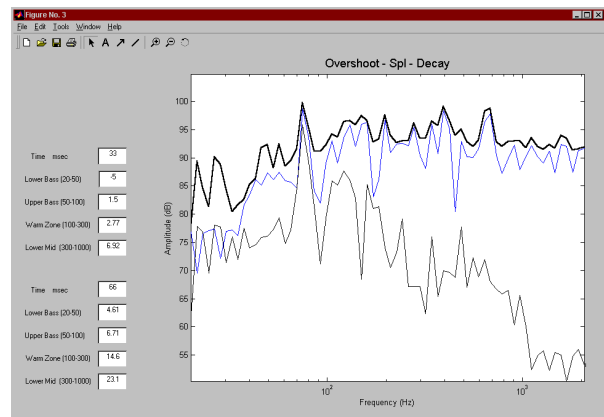


Fig. 17 – AQT2 Analysis of the Qusted loudspeaker

In fig. 17 it is clear how, at the room's resonance frequency of 74 Hz, the decayed level after 66ms substantially coincides with the steady-state level, and comes very close to the overshoot level.

Looking at the overshoot frequency response, this appears to be much flatter than the steady-state response. Furthermore, the

overshoot response is very close to the anechoic response of this loudspeaker, as measured in the ASK's anechoic chamber. What is important to note here, is that the subjective response is corresponding with this flatness, the listeners appreciate the sound of this loudspeaker, and perceive it as inherently flat and uncolored (as it is in the anechoic room), although the actual behavior in the listening room seems to indicate a severe frequency-response alteration.

This means that our ears sense the overshoot response, not the steady-state one. And thus it is better to use the overshoot curve for describing the frequency response of a sound system (loudspeaker plus small room).

This was confirmed from an informal listening test: the Qusted loudspeakers were digitally equalized, making use of two different sets of FIR coefficients: in the first case, the steady-state response was inverted, in the second one the overshoot response was used for computing the equalization curve. 5 subjects were asked simply to choose what equalization they preferred, and all them decided without doubt for the second one, designed by flattening the overshoot response.

Finally, a graph of the modulation depth (called also here "articulation" for compatibility with the original AQT software) is created, as reported in fig. 18. On the graph, the difference between the steady-state level and the decay level after 33ms and 66ms ms are plotted: the lower curve is always the 33ms one, and it is significant only for car audio systems, whilst the upper curve is the 66ms one.

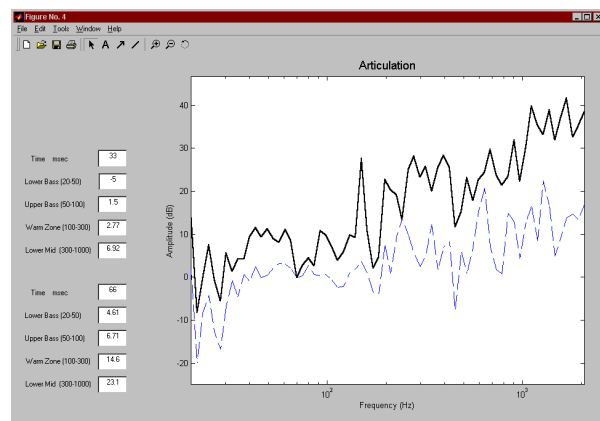


Fig. 18 – "articulation" vs. frequency

It can be seen that, at some frequency, the "articulation" is negative, as the level at 66ms (or more often at 33ms) is greater than the steady state level. This effect occurs typically when strong overshoot are present at the burst sides, and suggests again that a better metrics could be obtained measuring the modulation depth starting from the overshoot level instead of from the steady state level.

## CONCLUSIONS

The main goal of this paper was to describe three novel, similar objective qualification methods suitable for the analysis of loudspeakers installed in small rooms or cavities.

The original AQT method by Adami & Liberatore was described first [6], followed by the two more advanced implementations made by the authors of this paper. The final, revised method, called AQT2, is being proposed here as a new way of deriving useful information from measured impulse responses, showing the strict interrelationship occurring in these small spaces between frequency-domain and time-domain concepts.

## ACKNOWLEDGEMENTS

David Johnston, author of CoolEditPro [12], is acknowledged for his excellent software, which was kindly made available free for this research.

The loudspeakers employed for the tests in the ASK listening room were kindly provided free of charge from Audio Link srl, Alberi, Parma, Italy.

This work was supported through a research convention between ASK Industries, Reggio Emilia, Italy and the University of Parma, co-funded by the Italian Ministry for University and Research (MURST) under the grant MURST-98 #9809323883-007.

## References

- [1] A. Farina and F. Righini, "Software implementation of an MLS analyzer, with tools for convolution, auralization and inverse filtering", Preprint of the AES 103th Convention, New York, September 1997.
- [2] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", Preprint of the AES 108th Convention, Paris, February 2000.
- [3] A. Farina, E. Ugolotti, "Subjective comparison of different car audio systems by the auralization technique", Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.
- [4] E. Granier, M. Kleiner et al., 'Experimental auralization of car audio installations' Journal of the Audio Engineering Society, vol. 44, n. 10, 1996 October, pp.835-849.
- [5] A. Farina, E. Ugolotti, "Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications", 16<sup>th</sup> AES Conference, Rovaniemi (Finland) 12-14 April 1999.
- [6] I. Adami, F. Liberatore, "La messa a punto del sistema Diffusori-Ambiente" – Acustica Applicata srl, Via Roma 79, Galliciano – Lucca - Italy ([fabiolib@skylink.it](mailto:fabiolib@skylink.it)).
- [7] D. Zingani, I. Adami, F. Liberatore – Sound Analyzer v. 4.0a - Acustica Applicata srl, Via Roma 79, Galliciano – Lucca, Italy.
- [8] A.M. Noxon "the Music Articulation Test Tone (MATT)" - Acoustic Sciences Corporation (ASC), 4275 West Fifth Avenue, Eugene, Oregon 97440 - U. S. A. <http://www.tubetrap.com/matt.htm>
- [9] E. Zwicker and H. Fastl, Psychoacoustics - Fact and Models, Springer-Verlag, Berlin, 1990
- [10] T. Houtgast, H.J.M. Steeneken, R. Plomp "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics", *Acustica*, 16, 60-72 (1980)
- [11] Rife, D.D., and Vanderkooy, J., "Transfer-function measurements with maximum length sequences", *J. Audio Eng. Soc.*, June, 1989, 37, pp. 419-444.
- [12] Johnston, D., "Cool Edit Pro v. 1.2A", [HTTP://www.syntrillium.com](http://www.syntrillium.com), 2000.

## Internet References

The Aurora plugins [1,2] for the CoolEdit program [12] can be downloaded from [HTTP://www.ramsete.com/aurora](http://www.ramsete.com/aurora).

The sound samples employed in this paper can be downloaded from: [HTTP://pcangelo.eng.unipr.it/public/AQT](http://pcangelo.eng.unipr.it/public/AQT).

More info on related research, and a comprehensive list of publications, can be found on [HTTP://pcfarina.eng.unipr.it](http://pcfarina.eng.unipr.it).