Audio Engineering Society

Convention Paper

Presented at the 140th Convention
2016 June 4–7 Paris, France

Perception of low frequency transient acoustic phenomena in small rooms for music

Ing. Lorenzo Rizzi, Suonoevita [1], Federico Ascari, Politecnico di Milano [2], Ing. Gabriele Ghelfi, Suonoevita [3], and Ing. Michele Ferroni, Politecnico di Milano [4]

[1] Suono e Vita, Lecco LC, 23900, Italy
rizzi@suonoevita.it

[2] Politecnico di Milano, Como CO, 22100, Italy
federico.ascari@hotmail.it

[3] Suono e Vita, Lecco LC, 23900, Italy
ghelfi@suonoevita.it

[4] Politecnico di Milano, Como CO, 22100, Italy
michele.ferroni.ge@gmail.com

ABSTRACT

Reducing the gap between analysis of low-frequency behavior of small rooms and actual perception, introducing the importance of transient energetic phenomena besides classic FFT steady state analysis. After a frequency and temporal domain analysis of real-world impulse responses of critical listening rooms, headphone tests were performed. Results show that, for short musical sounds, a new curve called “Overshoot Response” can be more useful than classic frequency response regarding the level perception. Furthermore, the perceived loss of definition after the convolution with R.I.R. is correlated with decaying time and two metrics that were defined “Room Slowness” and “Room Inertia”.

1. INTRODUCTION

Low-frequency behavior of small rooms has been studied in theory, but the psychoacoustic confirmation of such analyses is minimal in the body of literature [1]. This research aims at reducing the gap between theoretical analysis and actual perception, also introducing new insight on the importance of transient energetic phenomena rather than classic FFT, steady state analysis.

2. LOW FREQUENCY IN SMALL ROOMS

2.1. Room modes and classic decay theory

The sound field in small rooms is to be considered deterministic at low frequencies. Schroeder [2] defined a theoretical critical frequency above which Sabine’s assumptions can be used. Real small rooms for music have short reverberation times and the furniture’s sound diffraction creates an intermediate region at the critical frequency. The authors’ experience defines a domain between 30 Hz and 300 Hz where modal influence is important for listening and playing music.

The phenomenon of resonant modes in small rooms is well-known: at frequencies whose wavelength is related with the room's dimensions, a standing wave is formed. These define the room's frequency response and make it highly dependent on the speaker's and receiver placement. Considering a shoe-box shaped room, the resonant frequencies can be computed by:

$$f = \tfrac{\omega}{2\pi} = \tfrac{kc}{2\pi} = \tfrac{c}{2}\sqrt{\left(\tfrac{n_x}{l_x}\right)^2 + \left(\tfrac{n_y}{l_y}\right)^2 + \left(\tfrac{n_z}{l_z}\right)^2} \tag{1}$$

Where c is the speed of sound, l_x, l_y and l_z are the room's dimensions and n_x, n_y and n_z are non-negative integers. Instead, if the room's dimensions are irregular, resonant frequencies are more difficult to compute, but still exist [3]. An example of such frequencies shown on the frequency response plot of a room can be seen in [4].

Classic theory can be simpler for a single mode [5]: when emitting a constant sound in a room, the sound pressure at that resonant frequency will build up until the magnitude of its RMS value equals:

$$|p_n| = \frac{K}{\delta_n} \tag{2}$$

where K is the generic source constant determined by the strength and location of the source and by the volume of the room and δ_n is the damping constant determined principally by the amount of absorption and by the volume of the room. Instead, if the sound is not centered on a room mode, the magnitude of the sound pressure at regime is given by:

$$|p_n| = \frac{2K\omega}{\sqrt{4\omega_n{}^2 k_n{}^2 + \left(\omega^2 - \omega_n{}^2\right)^2}} \tag{3}$$

where ω is the angular driving frequency and ω_n is the angular normal frequency.

If only one mode of vibration is excited, the decay is described by:

$$p_n = \frac{K}{\delta_n} e^{-\delta_n t} \cos(\omega_n t) \tag{4}$$

From this equation, the time required for the pressure to drop by 60 dB is:

$$T = \frac{6.91}{\delta_n} \tag{5}$$

Accurate modal decaying time measurements [6] show that the decaying time is never constant in frequency. Real frequency responses show an interaction between room modes, better synthesized by Green function [3]:

$$P_\omega(r) = \rho_0 c^2 \omega Q \sum_n \frac{p_n(r)p_n(r_0)}{\left[2\delta_n\omega_n + i\left(\omega^2 - \omega_n^2\right)\right]K_n} \tag{6}$$

This formula describes the transfer function of the room between points r and r_0 where Q is the source strength. Each term of the sum represents a resonance of the room, whose corresponding frequencies, given by $f_n = \omega_n / 2\pi$, are the eigen-frequencies and δ_n is the mode's damping constant.

2.2. Past research on room modes psychoacoustic perception

Early investigations on the subject found that resonance modes detection generally worsens as frequency decreases for steady state signals [7], and that temporal features such as transients highly impact the perception threshold of these phenomena. The perception thresholds of Decay Times have been found to grow with decreasing frequency below 100 Hz [8], suggesting that the reduction of modal decays below a certain threshold might be unnecessary, and a detection threshold of 16 for the Q factor has been found in [9]. More recent research on headphone tests [1] confirms such hypotheses, highlighting the importance of the content of the stimuli in addressing the problem, and defining a decay time threshold for the perception of room modes in frequency. The threshold is higher for artificial stimuli than for the tested music tracks. For the last category of signals, detection of modal effects is caused by both temporal and tonal effects.

The previous results were found by using synthetic room models and non-musical test sounds, when musical tracks were used they were generic recorded music [1] and [9]. The present research has its origins in the analysis of real small rooms and the influence on musical low frequency sounds perception: its aim is to analyze the effect of variables such as the spectrum content and the sound time envelope with regards to the perceived level and quality degradation.

3. FIRST GROUP OF LISTENING TESTS

Initially, the impulse responses of eight real rooms for music with volume between 30 and 56 m^3, measured by SuonoeVita, were analyzed [10]. Problematic frequencies were defined in the frequency response and decay times analysis, hence test sounds that excited those frequencies were developed (sounds whose fundamental frequency was placed on a peak or valley in the frequency response).

Since the aim was to find a psychoacoustic relation with music perception, sounds needed to be generated by musical instruments instead of being pure tones or recorded music. Synthetic kick drums and sampled DI box recorded bass sounds were used in order to avoid the influence of the room they would have been recorded in.

The sounds were then convolved with the RIRs, and the test asked simple questions regarding the perceived volume, sound quality, level of degradation of two or a sequence of test sounds.

Two generic tests were performed with 20 listeners each, most of them with musical background, on headphones, in order to avoid adding the influence of a listening room. Headphone low frequency tests are not expected to generate results that differ from loudspeaker tests [11] and [1]. All test conditions were identical for all listeners (playback devices and volume, test sounds, silent environment). Both tests were composed of about 20 questions, and included sounds with different spectral content (kick drum hits, bass notes, musical excerpts), different durations, different number of repetitions (sequences such as non-canonic musical scales modified to excite the eigen frequencies); sounds were either dry or convolved with different RIR. The full results can be found in [12].

The results of the first generic test can be summarized as follows:

- The test confirms the level perception to be in accordance with the loudness curves when sounds are dry, not convolved with RIRs, since 85% of testers perceived the highest note to be louder.

- After the convolution with RIRs, the presence of resonances can overcome the loudness curves importance. When the resonance was on the highest pitched note, the percentage of listeners that perceived it as loudest rose to 95%. Most subjects would perceive a higher level for a low-pitched note, if a strong resonance was present at that sound's fundamental frequency; in the test, it happened for 60 % of listeners in this scenario, whereas the remaining 40% stated that the loudest note was the one with highest pitch (which was, actually, the second loudest note).

- As expected the peak level is not an indicator of the perceived loudness of a sound. Resonance modes can cause a change in the time envelope, influencing its level and coloring its spectral content even with a lower peak.

- All subjects perceived a worsening of listening quality after the convolution. However, not all rooms created the same level of perceived degradation.

- The preference of a listening condition appears to be correlated with the perception of the attack of sound, which is modified by the presence of a resonance resulting in an envelope alteration and a timbric change. 90% of listeners indicated that the best playback quality was the one created by a room whose frequency response was quite flat, and whose decay times were below the perception threshold proposed by Avis [9].

After analyzing the results of the first test, a second test was developed two months later with new sounds and performed in order to gain more information. The main results are as follows:

- 50 % of subjects stated that decays were really disturbing when the room featured decays over the literature threshold [9], whereas 0% found them disturbing in a room with decays under the threshold.

- On short sounds like short bass notes and kick hits, the level differences are less perceived.

- The degree of certainty in subjects' answers regarding the effect of resonance modes on the perceived level grows for longer notes.

- Also for more complex sounds like musical excerpts, the room with less problems due to resonance modes was still preferred by 95 % of listeners. No significant change in the perception of resonance modes happened when high frequency instruments were added on top of the same sequence.

- By convolving sounds and musical excerpts with the two “best” environments preferences were unclear.

Furthermore, both tests were aimed at developing a vocabulary of adjectives that were descriptive of the perceived effect of resonance modes as done at Salford University [13]. In the first test, listeners were asked to describe some convolved sounds, and the resulting words were validated in the second test, where listeners had to match a sound with the terms that were most descriptive. The most recurring terms were the Italian words for: “precise”, “definite”, “dry”, “clean” against “resonant”, “damped”, “dark”, “reverberant”, “fat”, “confused”, “dirty”, “distorted”, “boomy”. These adjectives basically divide rooms between those that introduce a high degradation in the sound, and those who do not. As stated later in this article, this phenomenon seem to be correlated with the temporal behavior of the rooms.

4. AQT METHODOLOGY

The Acoustic Quality Test (AQT) is a measurement technique that allows to inspect the temporal evolution of test tones inside an environment [14]. This method is an evolution of MATT by M. Noxon [15]. The algorithm was then further refined by Farina et al. [16], first by creating its virtual counterpart, and then by developing the AQT 2 method, which offers more solid results.

The AQT 2 algorithm creates short sine bursts at increasing frequencies, and synthetically convolves each one of them separately with the environment’s impulse response. The output of each convolution (which, all together, create the 3D EFT plot) is the temporal evolution of each frequency in the room at the measurement positions. The envelope of these signals will be referred to as “Response Envelope” in the following: it allows to see the transient behavior for short sounds which is important for level and tone perception.

4.1. AQT analysis on FFT peaks and valleys

When analyzing a room using this algorithm, three main different behavior can be found, as Fig. 1 shows:

- On the peaks of the FFT curve, Response Envelopes are quite slow both in rising to and decaying from the steady state; for short tone bursts, they often fail to reach the actual steady state value (Fig. 1a, yellow line): these frequencies will be referred to in the following as “slow” frequencies.

- In some rooms, Response Envelopes of peaks reach their steady state also for short bursts (Fig. 1b) This behaviour defines “fast” frequencies and it is related to good acoustic correction.

- On the valleys of the FFT curve, Response Envelopes are faster both in rising to and decaying from the steady state, and they show an overshoot behavior, meaning that there is a peak higher than the steady state level in either or both the initial and final part of the Response Envelope (Fig. 1c). The lower steady state value, instead, is caused by the interference between direct and reflected field in the definition of the standing wave anti-node [16]. These frequencies will also be referred to as “fast frequencies”.

- On intermediate frequencies, the behavior changes gradually between the one on peaks and on valleys.

4.2. Steady State Response and Overshoot Response

Two useful curves can be computed by the AQT 2 algorithm: the “Steady State Response”, which shows the value reached by each frequency at the end of the burst (it equals the FFT curves for long tones) and the “Overshoot Response”, which shows the maximum value of the Response Envelope at each frequency. Therefore, the Overshoot Response shows the overshoot amplitude on valleys, and the maximum value reached on the peaks, making it greater than, or equal to, the Steady State Response at all frequencies. Both curves depend on the duration of the test sounds used in the simulation.

It is clear, therefore, that for short sounds the classic Frequency Response fails to describe the amplitude of the Response Envelopes (as hinted from the first tests’ results), because by definition it shows the actual steady state value at all frequencies, but short tones do not

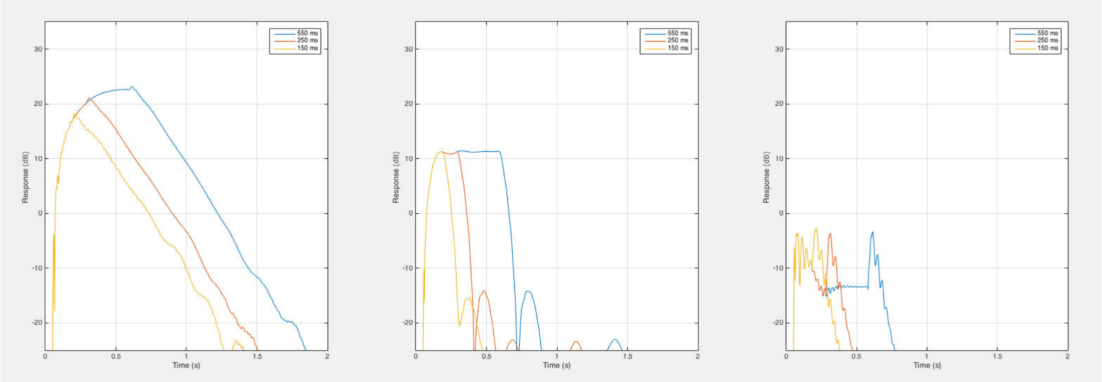

Fig. 1 – "Slow" frequencies on FFT peaks, "Fast" frequencies on FFT peaks, "Fast" frequencies and Overshoot Behavior on FFT valleys

always reach their steady state. One of the aims of this research was inspecting if the psychoacoustic perception of short sounds follows the classic Frequency Response or if the Overshoot Response is more significative as hypothized by Farina [16].

4.3. Analogy with Higher Order Systems

Response Envelope's rising behavior on peaks and valleys can be related to the step response of higher order systems, which depends on the natural frequency and damping ratio of the system to be controlled.

In simplistic terms, each term of Green sum (Eq. 6) can be described through classic control theory with a second order system. The output of such systems, in Laplace domain, when the input is a unitary step, is:

$$Y(s) = \frac{{\omega_n}^2}{s(s^2 + 2\varsigma\omega_n s + {\omega_n}^2)} \tag{7}$$

where 1/s is the Laplace transform of the unitary step, ω_n is the natural frequency and ζ is the damping factor. This is related to Green function through a different definition of the damping parameter.

In order to solve the anti-transformation, three different cases have to be analyzed depending on the parameter ζ, which causes the poles to be in different positions, creating different temporal responses. Figure 4 shows these cases, highlighting the effect of the damping factor ζ on the shape of the temporal response of this simplified system. As a matter of fact, the phenomenon of overshoots in the valleys of the frequency response of a real room confirms that the system is underdamped (0<ζ<1), while, on peaks, the system follows the overdamped case (ζ>1) and the envelope is slow in reaching its steady state.

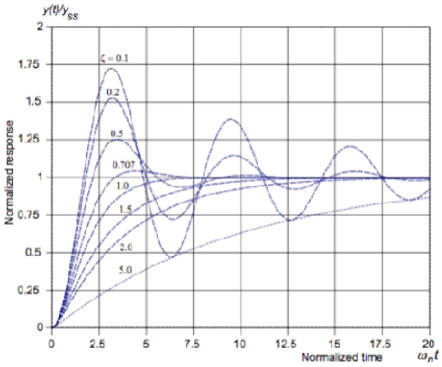

Fig. 2 – Step Response of higher order systems

More research should be made to better identify the nature of the damping factor in relation to the acoustic complex impedance of the room boundaries.

4.4. AQT Parameters Testing

The algorithm parameters were tuned to fit the needs of this research. In particular, all analyses have been carried out for three different durations of bursts: 150 ms (for which most peaks' Response Envelopes are not able to reach their steady state), 550 ms and an intermediate value of 250 milliseconds. These values were chosen evaluating common notes durations such as

125 ms (1/16th at 120 bpm) and 500 ms (1/4th at 120 bpm).

A preliminary analysis was also conducted to decide the most appropriate fade-in and fade-out durations for the tone bursts in the AQT simulation, as the overshoot behavior is influenced by this variable (as expected the overshoot behavior is higher when the fade-in is shorter). The test burst fade-in and fade-out values were chosen by inspecting real low frequency musical sounds (kick drum, bass). A fade-in value of 5 ms was chosen in both cases; a fade-out value of 20 ms was chosen to simulate bass envelopes, while the fade-out curve started just after the attack portion in order to simulate kick sounds, which are non sustained.

AQT simulations were done with both types of envelopes. The results are very similar with respect to the overshoot response but quantities defined by the steady state value (such as the frequency response, steady state response, Slowness and Inertia parameters) are not completely well defined for impulsive sounds since they do not reach a steady-state.

It will be showed that the temporal parameters derived by the simulation with sustained sounds still prove meaningful in the perception of non-sustained sounds. In this paper, only plots regarding sustained sounds are showed for space constraints, but this difference should be kept in mind. With these parameters, 8 rooms were analyzed in the frequency range between 20 and 300 Hz to inspect problematic frequencies.

4.5. AQT methodology improvements and new parameters definition

AQT 2 was further modified by the authors implementing more functionalities, adding decay time computation and further temporal behavior analysis, advanced overshoot quantity analysis, waterfall plot.

For all rooms, the Overshoot Quantity (Fig. 3, orange line) has been computed as the difference between the Steady State Response value and the Overshoot Response value at that frequency, and plotted on the frequency response, confirming that the overshoot quantity is maximum in presence of a valley, and is minimum on peaks.

Decay times were computed for all frequencies with a Schroeder Backward Integration using the slope over a fall in amplitude of 20 dB from the steady state value and multiplied by 3 to estimate a 60 dB drop. This algorithm outputs unnaturally high values when the steady state is particularly close to the noise floor (on valleys).

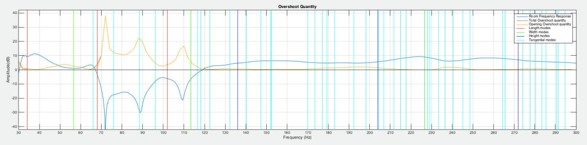

Fig. 3 - Overshoot Quantity in Room PRZ

4.5.1. Room Inertia

In order to overcome the decay algorithm limitation, a parameter called "Room Inertia" was introduced by the authors. This curve shows, for each frequency, the time passing between the end of the tone burst and the moment when the Response Envelope reaches a fixed, "target" value which is initially set to be the minimum value in the frequency response between 30 and 300 Hz, minus 1 dB in order to ensure that the R.I. value is always greater than zero. This curve allows to see intuitively which frequencies decay more rapidly.

4.5.2. Room Slowness

Similarly, a parameter called "Room Slowness" was introduced. This curve shows, for each frequency, the time passing between the instant when the test tone is started, and the moment when the Response Envelope reaches the steady state value minus a fixed value. Four different computations were done by setting the threshold at different values (0.1, 2, 6 and 10 dB) in order to obtain a different precision: the more the threshold is close to the steady state value, the more this measure takes into account the last part of the rise, which is usually slower; when the threshold is quite low, the measure mainly accounts for the initial slope of the Response Envelope. When combined together, they offer an intuitive view of how each frequency grows with time. As expected frequencies that have high Slowness values also have high Inertia values, and are mainly the, already defined, "slow frequencies". Instead, "fast frequencies" show very little values for both parameters.

4.5.3. Room rating

In order to characterize each room with just one global value for each temporal parameter, the mean of the vectors Decay Time, Room Slowness and Room Inertia was computed between 30 and 300 Hz for each room, rating each one of them and creating rankings with respect to each parameter. Rooms with high values for these parameters are referred in the following as “slow rooms”, whereas “fast rooms” will be used as the opposite. The aim of the fourth test was to inspect the correlation between these values and the perceived loss of precision and definition after the convolution with each room’s IR.

5. ROOM ANALYSIS RESULTS

AQT simulations were performed on all eight rooms between 20 and 300 Hz with three different test burst durations (150, 250 and 550 ms) and with pure tones having the two different envelopes described earlier, simulating both sustained and impulsive sounds. As already stated, results will be showed only for sustained, 550 ms tone bursts.

The results are in accordance to the principles introduced earlier, showing that there is strong correlation between peaks and “slow frequencies”, and valleys and “fast frequencies” with high overshoots. It is important to highlight, however, that the amplitude of the overshoot and the slowness of the Response Envelope are not always proportional to the amplitude of the peak or valley in the frequency response, since interference between neighboring room modes can take place. Also, different rooms have different slowness of the Response Envelopes in reaching their steady state value on peaks and this is correlated with their boundary building materials and acoustic conditions. Because of space constraints, this sections shows Steady State, Overshoot Response and Temporal Parameters for two of the most interesting rooms, named room SNT and room PRZ, for sustained test bursts of 550 milliseconds.

5.1. Room SNT

Room SNT is a 46.7 m^3, non-symmetric listening room with tilted roof, single gypsum board surfaces and no acoustic treatment. It is an example of a “slow” room, because on the frequency response’s peaks, Response Envelopes grow and decay slowly, returning high values for Room Slowness, Inertia (Fig. 5) and Decay Time. As an example, the peak at 70Hz does not reach its steady state value with short bursts. Valleys at 64, 128 and 230 Hz, instead, show very high overshoots values.

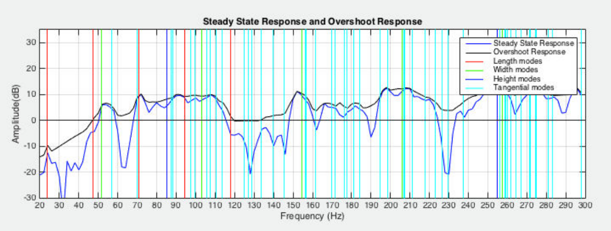

Fig. 4 - Steady State and Overshoot Response, Room SNT

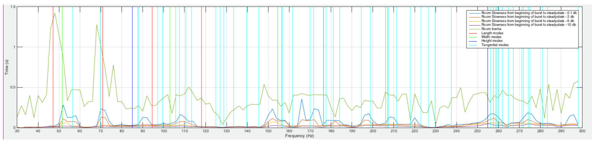

Fig. 5 - Slowness and Inertia, Room SNT

5.2. Room PRZ

Room PRZ is a 38.4 m^3, symmetric but irregular, mixing room with gypsum boards surfaces, heavily treated with sound absorbing mats. This room is quite fast at reacting to almost all of its frequencies. Frequency response peaks at 38 (its Response Envelope is visible in fig. 2), 64, 154 Hz reach their steady state value even with 150 msec bursts. The room also shows a large frequency valley in the low-frequency area due to wrong room proportions. However, the Overshoot Response in the same area is higher (Fig. 6), indicating high overshoots. With very low values of Slowness, Inertia (Fig. 7) and Decay Time parameters, this is used as an example of a “fast” room.

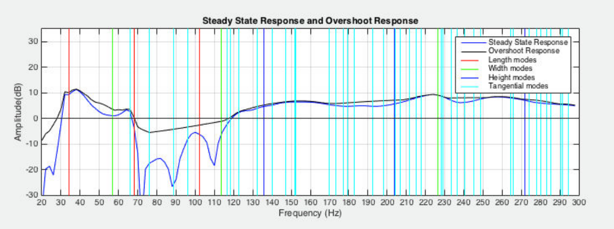

Fig. 6 – Steady State and Overshoot Response, Room PRZ

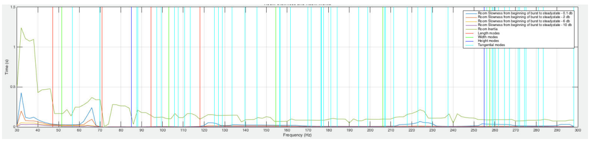

Fig. 7 - Slowness and Inertia, Room PRZ

5.3. Ranking of all rooms with respect to temporal parameters

The global values for the temporal parameters (computed as the average between 30 and 300 Hz) for each room are:

Room	Slowness	Decay Time	Room Inertia
CN	0.0077 s	0.4779 s	0.1880 s
PRZ	0.0047 s	0.1682 s	0.1695 s
SGR	0.0054 s	0.4416 s	0.0935 s
GD	0.0054 s	0.5257 s	0.1951 s
SNT	0.0158 s	0.6497 s	0.3692 s
DrmA	0.0035 s	0.3779 s	0.1507 s
DrmB	0.0030 s	0.4511 s	0.1601 s
DrmReg	0.0060 s	0.2638 s	0.1486 s

Table 1 Temporal Parameters for all rooms

The four most interesting rooms were chosen according to their temporal parameters and spectral content: Room SNT, Room PRZ, Room CN, Room SGR. Room CN is a 30.6 m^3 untreated mixing room with tilted roof and masonry walls in which frequency response's peaks show "slow" behavior, generating high Slowness and Inertia values, but not as high as room SNT. Room SGR is a 43.1 m^3 treated symmetric mastering room with gypsum board surfaces, a "fast" temporal behavior, and a frequency response which is flatter than Room PRZ. This room will be used in order to study the psychoacoustic influence of room PRZ's large valley with high overshoots with respect to the level perception. Ranking these rooms according to these parameters, the results are (from lowest to highest value, meaning from "fastest" to "slowest" room with respect to that parameter):

	"fastest"			"slowest"
Decay	PRZ	SGR	CN	SNT
Inertia	SGR	PRZ	CN	SNT
Slowness	PRZ	SGR	CN	SNT

Table 2 Room rankings according to temporal parameters

For these rooms the single value rankings are quite similar, confirming the correlation between the three parameters and basically giving an overall idea of the impact of each room after the convolution with a sound. Of course, many more rooms should be studied. The relationship between these rankings and the perceived quality degradation after the convolution was inspected during test four.

6. FURTHER LISTENING TESTS

After analyzing all rooms with the enhanced AQT algorithm, other two psychoacoustic tests were performed with the same conditions as the previous ones, with 30 listeners each [17]. The following sections contains some of the most interesting questions and results of tests three and four. The file naming is: typeofsound_duration(msec)_fundamentalfrequency_roomIR. The field "duration" is missing for kick sound which are impulsive, and the field "roomIR" is present only when the sound is convolved. All sounds have their fundamental frequency at one of the problematic frequencies in the specific room and are convolved with the RIR of that same room in order to highlight that behavior.

6.1. Test three

Question: which sound in each couple is the most precise? which is the most resonant?

Audio Files	Most precise: 1st	Most precise: 2nd	Most resonant: 1st	Most resonant: 2nd
Kick_44_CN Kick_38_PRZ	16,67 %	83,33 %	96,67 %	3,33 %
Bass_150_38_PRZ Bass_150_44_CN	90 %	10 %	0 %	100 %

Table 3 Question results – Precision and resonance

Results show that, comparing sounds tuned at the slowest resonance peak of each room, listeners perceive as more precise the one convolved in the fastest room and as more resonant the one convolved in the slowest. Also, the percentage depends on the type of sound as the first tests hinted, and is generally higher regarding the resonant quality.

Question: from the first to the second sound, does the precision get better or worse?

Audio Files	Better	Same	Worse
Kick_44 - Kick_44_CN	0 %	6,67 %	93,33 %
Kick_38 - Kick_38_PRZ	16,67 %	10 %	73,33 %
Bass_150_44_CN – Bass_150_44	83,33 %	16,67 %	0 %
Bass_150_38 – Bass_150_38_PRZ	20 %	56,67 %	23,33 %
Bass_550_44 – Bass_550_44_CN	16,67 %	16,67 %	66,67 %
Bass_550_38_PRZ – Bass_550_38	43,33 %	40 %	16,67 %

Table 4 Question results – Precision degradation

The perceived precision changes when the sound is convolved, according to the type of sound and characteristics of the room: slow rooms produce a more audible precision degradation than fast rooms, and this phenomenon can be heard mostly on impulsive sounds. Some listeners even stated that the precision gets better after the convolution with “fast” rooms, probably referring to a slight tone coloration. With longer sounds, the precision seems to be more confused for the slow room and slightly better for the fast room. More testing should be done to address this conflict.

Question: is the note always the same? on a scale from 1 to 10, how certain are you?

Audio Files	Same notes	Different notes
Bass_030_72 Bass_030_72_PRZ (valley) Bass_030_72_DrmReg (peak)	50 % Avg. certainty 7,73	50 % Avg. certainty 6,47
Bass_150_72 Bass_150_72_PRZ (valley) Bass_150_72_DrmReg (peak)	83,33 % Avg. certainty 8,32	16,67 % Avg. certainty 8
Bass_550_72 Bass_550_72_PRZ (valley) Bass_550_72_DrmReg (peak)	90 % Avg. certainty 8,15	10 % Avg. certainty 6,63

Table 5 Question results – Pitch perception

While the exact mechanism of pitch perception is not completely known yet, listeners definitely struggle in discriminating the pitch of very short sounds. Whether these results arise from a physiological factor alone, or the presence of resonance modes or valleys is actually able to upset the pitch perception for very short sounds, should be object of further studies. Some listeners stated that, from the first to the third very short sound, the pitch was perceived as descending.

Question: do the two sounds have the same volume? If not, which one is quieter?

Audio Files	Same volume	1st is quieter	2nd is quieter
Bass_150_99_PRZ Bass_150_111_PRZ	80 %	20 %	0 %
Bass_150_99_PRZ_CUT Bass_150_111_PRZ_CUT	60 %	3,33 %	36,67 %
Bass_150_111_PRZ Bass_150_111_PRZ_CUT	20 %	6,67 %	73,33 %

Table 6 Question results – Overshoot perception

In room PRZ, 99 Hz is a peak and 111 Hz is the center of a valley with high overshoots. Therefore, these frequencies have similar Overshoot Response value and different Steady State value. The first question compares normal short bass notes, while the second features file in which the initial and final part of the sound have been manually cut, removing the initial and final part of both envelopes (the overshoots) but doing so on a slightly longer note so that the resulting sound has the same duration as before. More testers are correctly able to identify the note centered on the valley. The third question compares the same sound on the valley with and without overshoots. On a direct comparison, 73 % of testers perceives as quieter the one without overshoots. These results prove the importance of energetic transient phenomena, and therefore of the Overshoot Response, in the level perception of short notes.

Question: do the two sounds have the same volume? If not, which one is louder?

Audio Files	Same Volume	1st is louder	2nd is louder	3rd is louder
Bass_150_111_PRZ Bass_250_111_PRZ Bass_550_111_PRZ	80 %	0 %	6,67 %	13,33 %
Bass_150_44_CN Bass_250_44_CN Bass_550_44_CN	60 %	3,33 %	13,33 %	23,33 %

Table 7 Question results – Level perception

Sounds are generally perceived as having the same volume, even though the results hint at the fact that longer sounds (both on peaks and valleys) can be perceived as being slightly louder with increasing durations. This behavior was expected especially in the second question because that frequency is “slow” in reaching its steady state value.
After analyzing these results, the concepts and algorithms for Room Slowness and Inertia were developed and a fourth test was prepared.

6.2. Test four

Question: which sound in each couple is the most precise? which is the most resonant?

Audio Files	Most precise: 1st	Most precise: 2nd	Most resonant: 1st	Most resonant: 2nd
Kick_32_PRZ Kick_38_PRZ	63,3 %	36,7 %	76,7 %	23,3 %
Bass_550_32_PRZ Bass_550_38_PRZ	16,7 %	83,3 %	83,3 %	16,7 %

Table 8 Question results – Precision and resonance

Frequencies 32 and 38 Hz in Room PRZ show an unusual behavior: 32 Hz is a peak with lower amplitude but higher Slowness and Inertia values than 38 Hz.

Results confirm that "precision" is not the best term to inspect the perceived degradation, because it creates discordant results with different types of sound. The word "resonant" offers more consistent results, indicating as the most resonant the sound whose fundamental has higher Slowness and Inertia values. It may also be the case that the difference on a single frequency is more hidden by the complex spectral content.

Question: do the two sounds have the same volume? If not, which one is louder?

Audio Files	Same volume	1st is quieter	2nd is quieter
Kick_38_SGR Kick_32_SGR	66,7 %	10 %	23,3 %
Bass_150_38_SGR Bass_150_32_SGR	76,7 %	0 %	23,3 %
Bass_550_32_SGR Bass_550_38_SGR	46,7 %	53,3 %	0 %
Bass_150_38_SGR_CUT Bass_150_32_SGR_CUT	56,7 %	10 %	33,3 %
Bass_150_32_SGR Bass_150_32_SGR_CUT	50 %	0 %	50 %
PureTone_550_32_SGR PureTone_550_38_SGR	6,6 %	80 %	13,3 %

Table 9 Question results – Overshoot perception

In room SGR, 32 Hz is a valley with strong overshoots and 38 Hz is a peak. Files of question four and five of this series were developed in the same way as before, manually removing the overshoot portion. The results confirm that the perception of overshoots is important and plays a role in the level perception. Also, when notes are short a level difference is difficult to hear, but for longer notes the low steady state value of valleys is exposed and the percentage of listeners who perceive the sound as having lower volume grows. Last question used pure tones, indicating that when a single frequency is affected, most people perceive a difference, indicating that masking occurs with sounds having a complex spectrum. Also, some people said that the first audio file had two peaks in its envelope, most probably referring to the overshoots.

Question: rank the sounds from the less precise to the most precise. Sounds: Kick_52_SNT (A), Kick_44_CN (B), Kick_38_PRZ (C), Kick_38_SGR (D).

ADCB	ADBC	BCDA	CDBA	CDAB	CBDA	DCBA	DCAB
3,3 %	3,3 %	3,3 %	16,7%	3,3 %	6,6 %	50 %	13,3%

Table 10 Question results – Kick precision

The ranking created by 50% of listeners is the same as the one according to the Room Inertia parameter. The second most rated alternative, with 16.7% of preference, was the same as the one generated by Room Slowness and Decay Time parameters.

Question: you will hear two sounds. Think of the first one as having "no degradation" and rate the second one against the first on this scale with respect to its perceived degradation, where the lowest part stands for "extreme degradation" and vice versa.

The scale [17] was developed taking into consideration the discussion in [18], using an apparently continuous scale with a gradient from black to white, with slightly loose labels of sensory nature describing the amount of perceived degradation. Anchoring technique was used, comparing the convolved sample to the dry one. A scale from 1 to 100 was present under the gradient part in order to record a number to perform statistic analyses. All combinations of four room and four types of sounds, each one tuned to the slowest mode of the corresponding room, were used, a musical excerpt was used as well. Average scale results are as follows:

	Room PRZ	Room SGR	Room CN	Room SNT
Bass_150	84,4	72,57	44,37	20,07
Bass_550	74,99	75,23	55,52	33,5
Kick	69,95	53,5	29,53	23,37
Excerpt	73,97	66,8	37,87	22,7

Table 11 Question results – Perceived precision in rooms

Results show that the room that performed less degradation was Room PRZ, and the ranking was the one given by the variables Room Slowness and Decay Time. The files referring to the Kick hits were the same as the previous question, indicating a slight ambiguity in the results. More tests should be performed to understand the reason, but it is clear that "faster" rooms generally introduce less precision degradation than "slower" rooms.

A two-way ANOVA analysis was performed with these results. Dealing with real-world Impulse Responses, variables were not controllable independently, so the behavior was analyzed with respect to the overall room temporal behavior (generally low to high values of all temporal parameters Inertia, Slowness, Decay Time) and with respect to the type of sound.

With a confidence level of 0.05, p-value is 0.000667 for the variable "type of sound", 0 for the variable "temporal behavior", and 1 for the combined effect of both variables. This means that both type of sound and overall temporal behavior are significant to the test results, and that the interaction between the variables is not significant, as expected, meaning that specific levels of one variable do not alter the general trend of results.

Question: rate each listening condition on this scale following your personal preference (no comparison was used in this question and a scale similar to the previous one but with only labels at its ends was used).

	Room PRZ	Room SGR	Room CN	Room SNT
Excerpt	70,07	78,1	29,9	19,5

Table 12 Question results – Overall listening preference

A very interesting consideration appears if comparing these results to the ones in the previous question. While room PRZ was perceived as introducing less degradation, room SGR is generally the preferred listening condition. Room PRZ has lower Decay and Slowness values, but it has a large low-frequency valley in its frequency response, while room SGR is flatter. The hypothesis is that when sounds are long enough to enter the steady-state domain described by the frequency response, listeners are able to perceive the frequency response. Many listeners, in fact, stated that room PRZ was “colder” whereas room SGR was “warmer”. This would mean that temporal parameters can be used to describe the sound degradation, but not always the listening quality.

One last section asked listeners to describe four sounds using descriptive terms from the pool developed in tests 1 and 2, adding the possibility to enter new adjectives. No relevant new adjectives were introduced, and the most significant words describing sounds in “fast” rooms were “Defined”, “Precise”, “Dry”, “Clean”, and those describing sounds in “slow” rooms were “Reverberant”, “Boomy”, “Dark”, “Fat”. The word “resonant”, interestingly, was not between the most chosen ones, probably because of its difficult interpretation. However, some words were more descriptive of a timbric quality (fat, dark) or the presence of an effect (reverberant, clean). The authors think that the most meaningful terms to describe sounds in “fast” rooms are “defined” and “precise”, whereas “boomy” should be used, in conjunction with “resonant”, do describe sounds in “slow” rooms. The reader is reminded that these tests were performed in Italy, so the resulting Italian terms (respectively, “definito”, “preciso”, “rimbombante”, “risonante”) were translated from italian.

7. CONCLUSIONS

From the results, it is confirmed that the psychoacoustic perception of listeners is greatly affected by the presence of room modes, which can alter the sound level perception and the perceived sound quality. This is important in music production and in high quality music listening.

The tests demonstrated that the effect of room modes is clear when the sound fundamental frequency actually excites one of the room’s eigen frequencies. This demonstrates that also from a perceptual point of view it is wrong to use a generic octave or third octave band reverberation time describing room decay at low frequencies.

It appears that the Overshoot Response can be a more useful counterpart to the classic FFT curve, regarding the loudness perception of low frequency short sounds. When sounds are long enough to enter their steady state domain, the perception is closer to the FFT curve.

Furthermore, it is clear that listeners are able to perceive the degradation caused to a sound by the convolution with an impulse response, to a degree that is correlated with the temporal characteristics of the room (more evident for “slow” rooms), the type of sound (more evident for impulsive sounds), and the duration (more evident for short sounds). Time is clearly a critical variable both on transient evolution and its perception.

Listeners seem to be more sensible to changes in precision and resonance rather than to changes in loudness. Specifically, all rooms were ranked according to Room Inertia, Room Slowness and Decay Time. Testers were asked to rate the perceived loss of definition of four different types of sounds (short and long bass notes, kick samples, music excerpts) each one convolved with four rooms responses with different temporal characteristics. Results show that the room ranking with the same order as the one defined by the testers is the one given by Room Slowness and Decay Time. In a similar question focused only on kick drum samples, the outcome corresponded to the rank given by Room Inertia. Room Inertia and Slowness could open new possibilities in both psychoacoustics and room design in parallel to decay time.

Further research is under study by the authors to investigate these results in a real listening environment, employing more orthodox psychoacoustic tests without the use of convolution and headphones. More research is needed on the nature of the decay parameter and on the interaction between the resonance's damping constant and the room boundary impedance.

8. ACKNOWLEDGEMENTS

The authors would like to thank prof. Giuseppe Bertuccio from Politecnico di Milano.

9. REFERENCES

[1] B.M. Fazenda et al. – "Perceptual Thresholds for the effects of room modes as a function of modal decay", J. Acoustic Soc. Am. 137 (2015) pp. 1088 - 1098

[2] M.R. Schroeder – "Schroeder Frequency Revisited" In J.Acoustic Soc. Am. 99 (1996) pp. 3240-3241

[3] H. Kuttruff – "Room Acoustic", fifth edition, Focal Press (2008)

[4] L. Rizzi et al. – "Small studios with gypsum board sound insulation: a review of their room acoustics, details at the low frequencies" – Audio Engineering Society, 124th Convention, Amsterdam (2008)

[5] L. Beranek – "Acoustics", Acoustical Society of America (1996)

[6] L. Rizzi et al. – "Room Acoustic measurements in non Sabinian enclosures for music: echometry, modal analysis, sound decay analysis", Internoise 2010, Lisbon (2010)

[7] F.E. Toole et al. – "The modification of Timbre by resonances: perception and measurement", J. Aud. Eng. Soc. 36(3), 122-141 (1988)

[8] M. Karjalainen et al. – "Perception of Temporal Decay of low frequency room modes", 116th Audio Engineering Society Convention, paper 6083 (2004)

[9] M.R. Avis et al. – "Thresholds of Detection for Changes to the Q Factor of Low-Frequency Modes in Listening Environments", J. Audio Eng. Soc., Vol.55, No 7/8 (2007)

[10] L. Rizzi et al. – "Small Rooms Dedicated to Music: from Room Response Analysis to Acoustic Design", Audio Engineering Society, 140th Convention, Paris (2016)

[11] V. Koehl et al. – "Comparison of subjective assessments obtained from listening tests through headphones and loudspeaker setups", Audio Engineering Society, 131st Convention, New York (2011)

[12] M. Ferroni – "Evaluation and Psychoacoustic Validation of Techniques for the Analysis of Low Frequency Resonance Modes in Real Small Rooms", Master thesis, Politecnico di Milano

[13] M. Wankling et al. – "The Assessment of Low-Frequency Room Acoustic Parameters Using Descriptive Analysis", J. Audio Eng. Soc., Vol. 60, No 5 (2012)

[14] I. Adami, F. Liberatore – "La messa a punto del sistema Diffusori – Ambiente", Acustica Applicata srl, Lucca, Italia

[15] A.M. Noxon – "The Music Articulation Test Tone (MATT)", Acoustic Sciences Corporation, Oregon USA

[16] A. Farina et al – "AQT – A New Objective Measurement Of The Acoustical Quality of Sound Reproduction In Small Compartments", Audio Engineering Society, 110th Convention, Amsterdam (2001)

[17] F. Ascari – "DSP Analysis and Psychoacoustic Testing on the Perception of Low Frequency Transients and Quality Degradation in Small Rooms", Master thesis, Politecnico di Milano

[18] S. Zielinski – "On Some Biases Encountered in Modern Audio Quality Listening Tests – A Review", J. Audio Eng. Soc., Vol. 56, No.6 (2008)