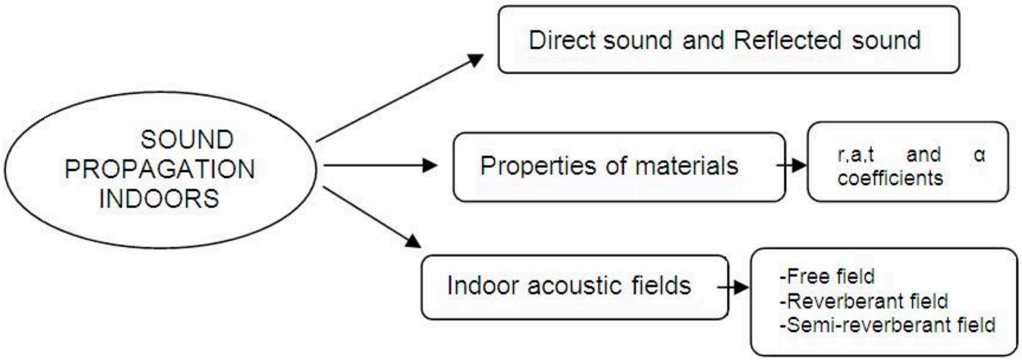
SOUND PROPAGATION INDOORS



Room acoustic is a very important topic: many places of our everyday life such as theatres, cinemas and classrooms are affected by this problem. In addition, noise effects in work places or buildings, where people is exposed to noise for the whole day, is not something taken for granted. For the first kind of indoor environment Sound Pressure Level is not the most relevant parameter. In fact, other parameters related to quality are used.

But how loud is a sound inside a room? What happens when we're placing a sound source in an enclosed space?

Direct and Reflected sound

In many kinds of enclosed spaces, the sound propagates in a peculiar way, very different from what happens outside (even if there are obstacles or sound barriers that can cause reflection). There still is a free field sound (direct waves) where waves travel directly from the source to the receiver. Of course if it is a point source the sound propagates with spherical waves; if a line one, with cylindrical waves.

But this is not the only sound. It is not even the most relevant one.

In most cases much more energy arrives to the receiver by a significant contribution due to room reflections: the sound bounces over the walls, floor etc. (reflected waves).

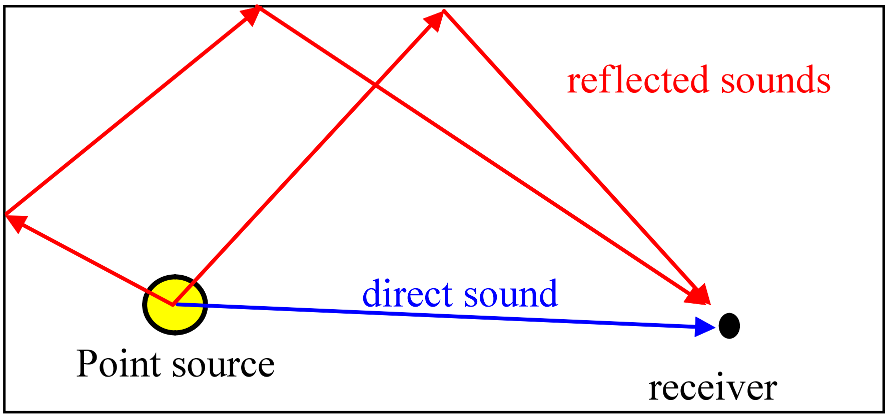
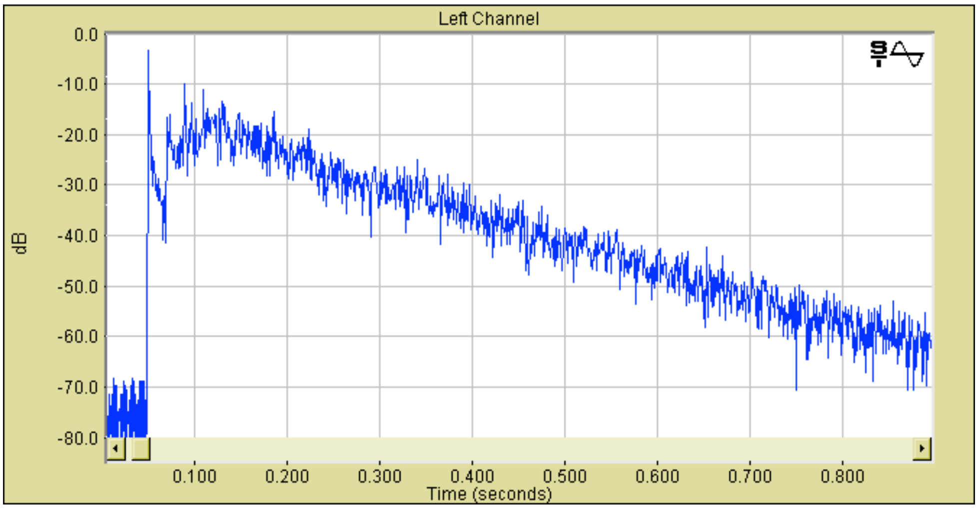


Fig.01 – different types of waves

In Fig.01, for example, sound is bouncing over the ceiling and a wall before reaching the receiver; surely there will be multiple reflections, also several times, with different and strange trajectories acting like a so-called "*acoustic billiard*".

So there is a significant contribution to the sound field due to reflected waves.

There is a strict relationship with light, which can reflect specularly (mirror) or diffusely (white plastered wall).



The chart above, known as "reflectogram", shows what happens inside a room by means of a so-called energetic impulse response. As a metter of fact the source has radiated an impulse (a hand clap for example) with a length of 2-3 milliseconds. Outdoor the response should have been different: just one peak, followed by silence, as before.

Inside a room, we do not ear only the direct sound: that is just the first sound we ear. After a while, reflections arrive from different ways that we're not able to know because pressure is an omnidirectional quantity, not a vector. Every peak in the chart is a discrete room reflection making longer and longer paths, arriving weaker and weaker.

So the sound pressure level inside a room can be given by two contributions: direct sound and reflected sound. In some cases one dominates on the other, but, in general, we have to keep both.

Properties of materials

The reflected energy depends on the nature of the surface: there are strongly reflective materials, so they reflect almost 100% of the energy. On the other hand, other materials are "acoustic absorbers", like a carpet. This last kind of materials can be really useful in order to reduce sound pressure level inside a room by reducing the amount of reflected energy.

Let's consider a wall for example (fig.02), which is separating two different rooms. When a sound source is placed on the left hand side room, the sound bounces over the separation wall and some energy is reflected, coming back in the source room. Some energy is dissipated inside wall's material. Another quantity of energy passes through the wall and the sound can be heard in the next room (like we can ear an aircraft flying over our house).

So energy is divided into three contributions: reflected energy, absorbed energy and transmitted energy.

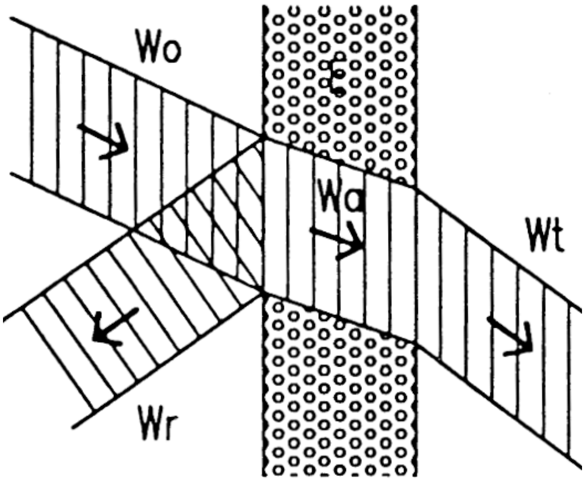


Fig.02 – sound energy subdivision over a wall

This also happens for light and electromagnetic waves (e.g. radio waves)

We can now define 3 numerical coefficients (*reflection, absorption and transmission coefficients)*

$$r=W_r/W_0\ ,\ a=W_a/W_0,\ t=W_t/W_0$$

Their sum must be one because of the energy conservation principle.

$$r + a + t =1$$

Of consequence, their value is bounded between 0 and 1.

$$0 < r,a,t < 1$$

A very reflecting surface (like a window glass) will have r=1; a,t=0. An open window is completely transmitting so t=1; r,a=0.

Acousticians, however, do not employ any of these three numbers. They only use the so called “apparent acoustic absorption coefficient” α, defined as

$$\alpha = 1 - r \qquad (1)$$

It expresses the energy which has not been reflected. That energy can be absorbed (a) or it can be transmitted (t). So it is possible to write α also as:

$$\alpha = a + t \qquad (2)$$

An open window has a unit value of α.

A perfect reflecting material has $\alpha = 0$.

All we need to know about the materials covering a room (ceiling, walls, floor etc.) is the apparent absorption coefficient α.

Indoor acoustic field

In an enclosed environment, the acoustic field can be of three different kinds: **free** (in which reflected energy is weak), **reverberant** (in which reflected energy is dominant) and **semi-reverberant** (intermediate case in which we have to consider both).

A field is defined as **free** when we are close to the source, where the direct energy component prevails: compared to it, the contribution of all the reflections becomes negligible. It exists in every room.

In this case, the field is the same as outdoors, and only depends on source distance and directivity, **Q.** Sound pressure level is

$$L_p = L_w + 10\log\left(\frac{Q}{4\pi d^2}\right)$$

In which **L_W** is the sound power level of the source, **Q** its directivity, and **d** is the distance between source and receiver. In a free field, the sound level decreases by 6 dB each time distance **d** doubles.

A field is said to be **reverberant** if the number of side wall reflections is so elevated that it creates a uniform acoustic field (even near the source).

The **equivalent acoustic absorption area** of the room is defined as:

$$A = \alpha S = \sum_i \alpha_i \cdot S_i \qquad (\mathrm{m}^2)$$

where α is the average absorption coefficient and S is the total interior surface area (floor, walls, ceiling, etc.).

In a purely-reverberant sound field, the Sound Pressure Level is:

$$L_p = L_w + 10\log\left(\frac{4}{A}\right)$$

Its value is the same everywhere in the room (diffuse field). A reverberant field may be obtained in so called *reverberant chambers,* equipped with diffusers and funny shapes sized, where the absorption coefficients of different materials are also measured. In real world it really doesn’t exist. For this reason this is just a theoretical case.

A field is said to be **semi-reverberant** when at every point the sound field is affected both by the free field and by the reverberant field.

In most normally sized rooms, we can suppose that the acoustic field is semi-reverberant. Sound pressure level is

$$L_p = L_w + 10\log\left(\frac{Q}{4\pi d^2} + \frac{4}{A}\right)$$

In a semi-reverberant acoustic field, the sound energy density at a point is therefore given by the sum of the direct and reflected acoustic fields.

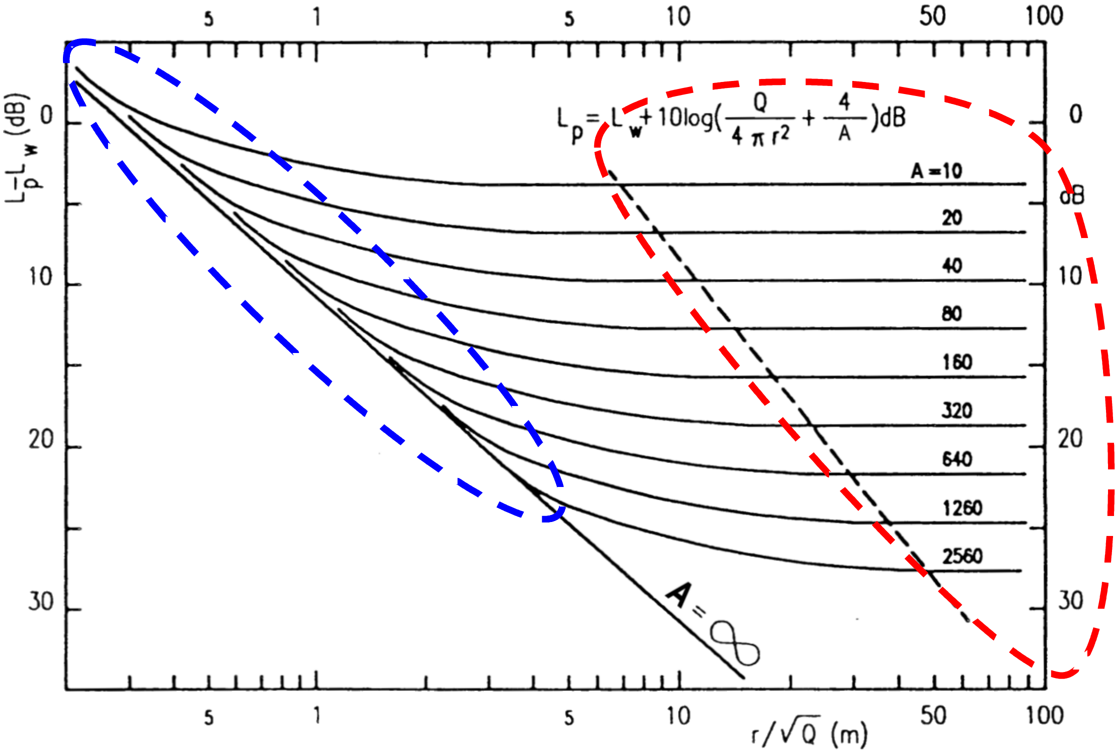


Fig.03 – semi-reverberant curves

In the graph above there are many curves: each one is plotted for a different value of the total room absorption A. The **blue zone** (A = ∞) represents the limit case for a free field while the **red** one marks a zone on which the acoustic field is practically reverberant. In between the two zones there is the semi-reverberant field where we should not neglect any of the two contributions.

If we take into account one single curve (for example the one for A=160m^2 as in Fig.03) we notice that reverberant sound level represented by the horizontal line and direct sound level represented by the sloped one, meet each other in a special point in which the two levels are the same. Known as critical distance, this point marks the separation between the two cases.

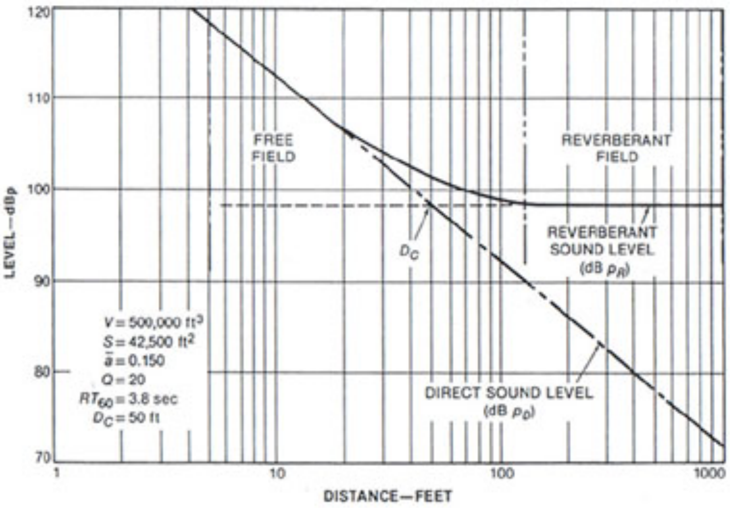


Fig.04 – critical distance point

We can now compute the value of critical distance by means of the semi-reverberant formula, where it is possible to notice two contributions.

$$L_p(d) = L_W + 10\cdot\lg\left[\underbrace{\frac{Q}{4\cdot\pi\cdot d^2}}_{\text{Direct sound}} + \underbrace{\frac{4}{\sum\alpha_i\cdot S_i}}_{\text{Reflected sound}}\right]$$

The two terms must be equal: then, solving the equation, we find the critical distance expression.

$$\frac{Q}{4\cdot\pi\cdot d^2} = \frac{4}{\alpha\cdot S} \qquad d_{cr} = \sqrt{\frac{Q\cdot\alpha\cdot S}{16\cdot\pi}}$$

We can notice that the critical distance is increased by the directivity of the source (very big loud speakers employed in sport arenas can project the sound very far); again, a large value of A=αS increases the critical distance. So in a room with small absorption there is a small critical distance, as in reverberant rooms.

At distances equal two or three times the critical distance, we can think that the sound field is fully reverberant. Hence, in a reverberant room, as we move only slightly away form the source, we find a perfectly diffuse sound field, with the same level everywhere.

Reverberation Time

The concept of reverberation time is central to acoustics and is considered the **mindstorm** of modern acoustics. It is related with the sound decay in time and was invented by professor Wallace Clement Sabine in 1902.

In general reverberation is the persistence of sound in a particular space after the original sound is produced.

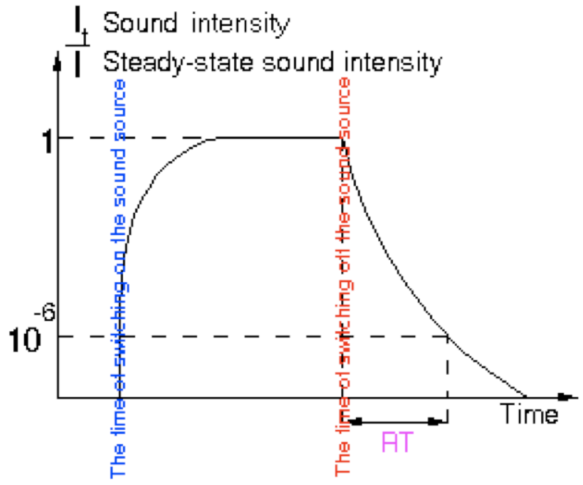
Sabine started his experiments using an organ as a sound source, a stopwatch and his ears; he measured the time from interruption of the source to inaudibility (approximately 60 dB) and found that the reverberation time is proportional to the dimensions of the room and inversely proportional to the amount of absorption present.

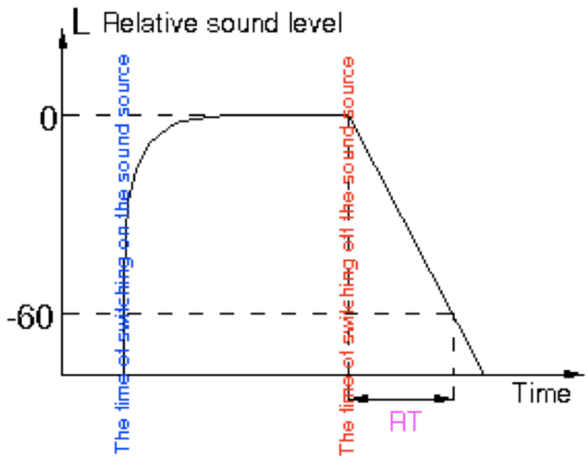
It is very important to notice that before the studies carried out by professor Sabine the acoustical evaluation of a room was only empirical.

Professor Sabine in 1902

Reverberation Time T_{60}

If we consider a room containing an active sound source and abruptly interrupt the emission of sound energy once it has reached a steady state we obtain the following results as shown in the charts below:





When we switch on the source the sound energy density increases until we reach an equilibrium in which as much power radiated by the source in one second is dissipated in one second.

After we have reached the steady state of sound energy density (in which the room is leaking sound at the boundaries but is continuously getting new sound from the source) we abruptly switch off the sound source.
We can easily verify how the decay is NOT linear.

But if we consider a chart of sound pressure level (dB) shown on the right we can observe that the decay is not exponential like before but linear; in other words the slope of the decay is CONSTANT.

The reverberation time is a strange way of expressing this slope; as a matter of fact normally the slope would be expressed by a quantity in dB/s. Instead the reverberation time was defined as how much time it takes for decaying a certain number of decibels (in particular 60 dB because in his experiment Sabine noticed that the distance between the sound produced by his organ and the background noise was 60 dB).

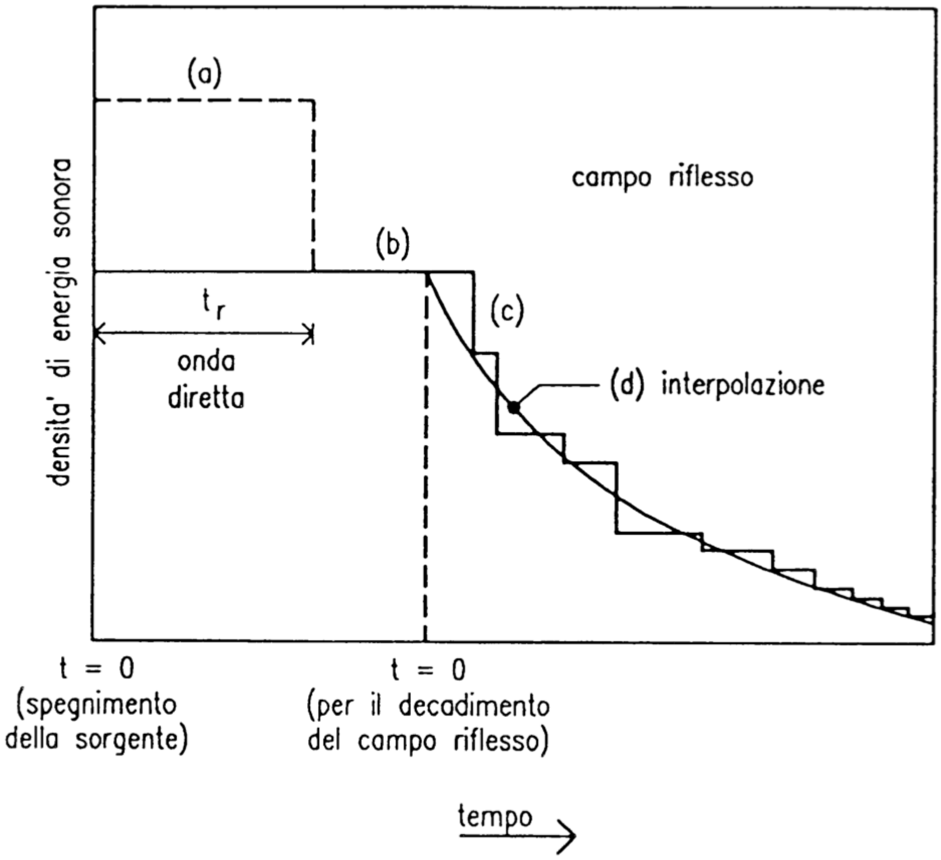
Reverberation time is therefore defined as the time necessary for the sound energy density to decrease to a millionth (60 dB) of the value it had before the source was switched off.

This is a formal definition, although isn't accurate in practical situations, because in a room the sound doesn't really decay of 60 dB.

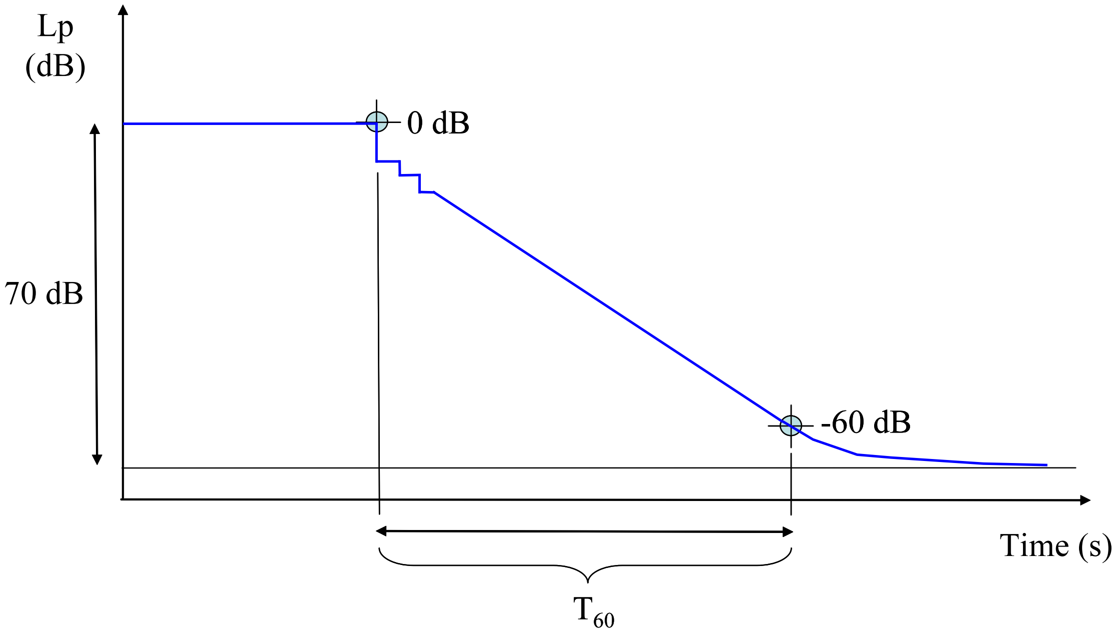
First of all the decay is not the continuous line which you would expect by the charts illustrated above.

In fact at the beginning for a certain number of decibels we can see some "gradons" that represent the loss of discrete contribution of the total energy every time a single discrete reflection goes away.

These gradons keep getting smaller in time but at the beginning they are very easy to distinguish.



A more realistic chart is shown below:



When the sound source is on we have a perfectly constant level, at a certain time the direct sound is interrupted so the sound level falls for a couple of dB.

Then for a short time the sound is still constant until the first reflection goes away and so on (sound reflections become smaller and smaller). After a while we see a decay going down with a certain slope. The curve in the end asymptotically approaches the background noise that is always present in a room.

The current international standard ISO 3382 – 2011 requires to compute not a reverberation time over a decay over 60 dB but recommends to measure the time for a decay starting at -5dB (relative to the steady state level) and ending at -25 dB.

By multiplying this decay time for 3 you easily obtain the True Reverberation Time (T_{20}).

The standard also defines T_{30} , T_{10} (unused) and EDT (early decay time).

T_{30} measures the decay between -5 dB and -35 db, T_{10} between -5 dB and -15 dB and EDT between 0 and -10 dB.

T_{10} ,T_{20}, and T_{30} need to be multiplied by a number of times so that the decay is 60 dB.

If the decay is perfectly linear $T_{10} = T_{20} = T_{30}$

In addition, for the measurements to be correct the end point of the decay must be at least 15 dB above the background noise level. If this does not happen the measurement must be repeated by using a more powerful sound source or by reducing the background noise level.

SABINE’S FORMULA

Professor Sabine defined the concept of Reverberation Time in Boston in 1902. Not only did he explain how to measure it but he also found the way of predicting the reverberation time of a not existing room; in this way you can know in advance how the reverberation time will be depending also on the type of materials of the room.

If the environment is perfectly reverberant the value of the reverberation time is the same in all points and is:

$$T_{60} = 0.16 \cdot \frac{V}{\sum_i \alpha_i \cdot S_i}$$

where V is the volume of the environment. It is also possible to determine the **equivalent area of acoustic absorption** $\mathbf{A = \alpha S}$

If we reverse the formula we can directly obtain the **total acoustical absorption area**:

$$T_{60} = T_{20} = \frac{0.16 \cdot V}{\sum \alpha_i \cdot S_i} \quad \Rightarrow \quad A = \frac{0.16 \cdot V}{T_{60}}$$

Substituting in the critical distance formula:

$$\mathrm{d_{cr}} = \sqrt{\frac{\mathrm{Q}}{16 \cdot \pi} \cdot \frac{0.16 \cdot \mathrm{V}}{\mathrm{T_{60}}}} = \sqrt{\frac{\mathrm{Q}}{100 \cdot \pi} \cdot \frac{\mathrm{V}}{\mathrm{T_{60}}}}$$

As we can see the critical distance is directly proportional to the volume of the room and inversely proportional to the reverberation time.

Acoustical Parameters

The ISO 3382 Standard

This standard defines the acoustical parameters related to the analysis of the impulse response of a room and particularly is based on quality of the acoustic in terms of musical perception and speech intelligibility.

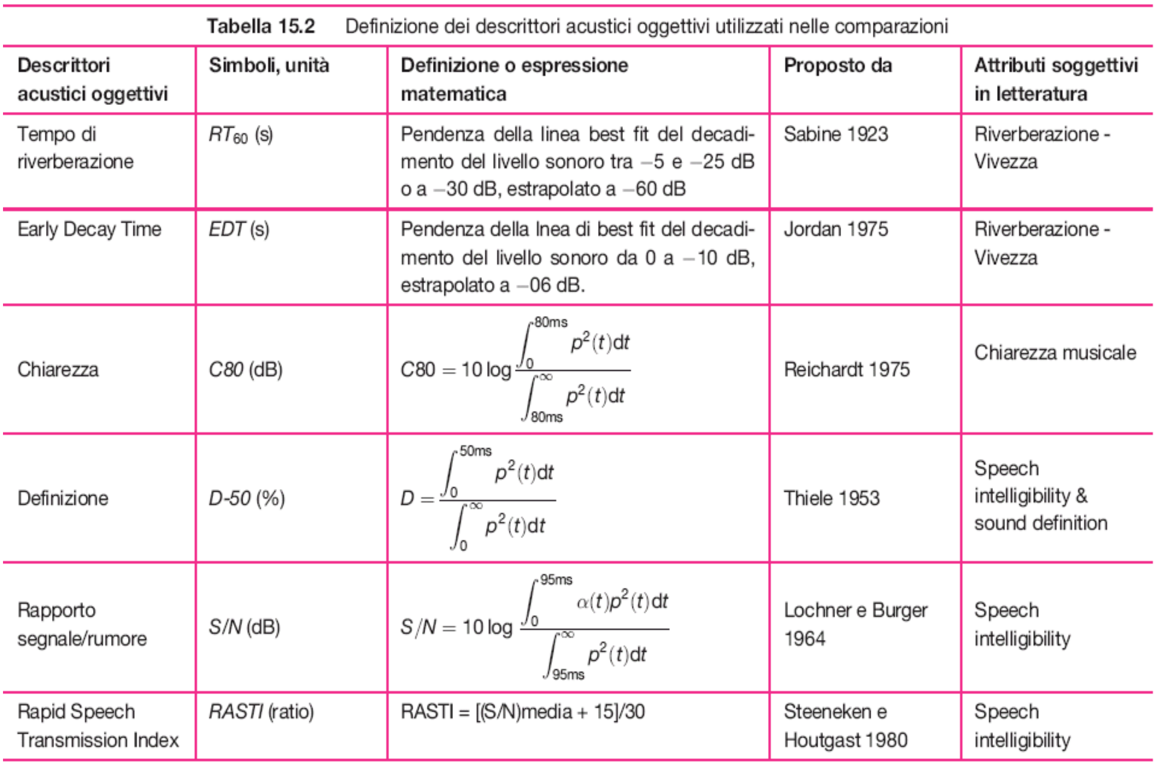


Tabella 15.2 Definizione dei descrittori acustici oggettivi utilizzati nelle comparazioni

Descrittori acustici oggettivi	**Simboli, unità**	**Definizione o espressione matematica**	**Proposto da**	**Attributi soggettivi in letteratura**
Tempo di riverberazione	RT_{60} (s)	Pendenza della linea best fit del decadimento del livello sonoro tra –5 e –25 dB o a –30 dB, estrapolato a –60 dB	Sabine 1923	Riverberazione - Vivezza
Early Decay Time	*EDT* (s)	Pendenza della lnea di best fit del decadimento del livello sonoro da 0 a –10 dB, estrapolato a –06 dB.	Jordan 1975	Riverberazione - Vivezza
Chiarezza	*C80* (dB)	$C80 = 10\log\frac{\int_0^{80ms} p^2(t)dt}{\int_{80ms}^{\infty} p^2(t)dt}$	Reichardt 1975	Chiarezza musicale
Definizione	*D-50* (%)	$D = \frac{\int_0^{50ms} p^2(t)dt}{\int_0^{\infty} p^2(t)dt}$	Thiele 1953	Speech intelligibility & sound definition
Rapporto segnale/rumore	*S/N* (dB)	$S/N = 10\log\frac{\int_0^{95ms} \alpha(t)p^2(t)dt}{\int_{95ms}^{\infty} p^2(t)dt}$	Lochner e Burger 1964	Speech intelligibility
Rapid Speech Transmission Index	*RASTI* (ratio)	RASTI = [(S/N)media + 15]/30	Steeneken e Houtgast 1980	Speech intelligibility

Fig.1 – ISO 3382 parameters.

From Impulse Response to Sound Decay

By the theoretical model:

Reverberation Time: time required for a sound decay of 60 dB.

But we are not having a source radiating a steady state sound with a constant power, we have an **impulsive source**.

We have a very short pulse (the direct sound radiated by the source), then a silence gap, then the first reflection (another pulse), then silence again, going to an increased temporal density of the reflections. This is not similar to the approximated linear sound level decay we can see switching off a steady state source.

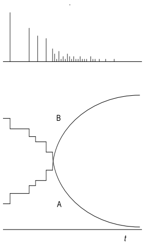


Fig. 2 – Pulsive Source and its reflections.

Indeed, we can’t apply the definition of reverberation time to the original impulse response.

We need to perform a time integration for transforming the response of the room to a pulse into the response of the room to a steady state sound switched off.

The steady state level produced by a source which continuos radiating the same power of the pulse radiated by our pulsive source is a **total time integral of the impulse response** h (or g), a total sound pressure level:

$$L_{p,tot} = \int_{o}^{\infty} h^2(t') \cdot dt' \qquad (1)$$

We can compute the decay subtracting each energy arrival of each reflection from the total energy. This total value is a **forward** or **upward running integral**:

$$L_{p,forw} = \int_{o}^{t} h^2(t') \cdot dt' \qquad (2)$$

The curve we are building is the total integral minus the running integral which comes back, thanks to the integral properties, to the backward integral (**Schroeder Backward Integral**):

$$L_{p,backw} = \int_{t}^{\infty} h^2(t') \cdot dt' \qquad (3)$$

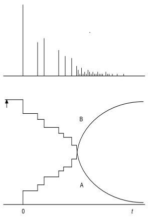


Fig. 3 – Computing sound decay: Total integral, running integral and Schroeder BW Integral.

Schroeder BW Integral transforms impulse respose into the decay of a stationary source, obtaining a very accurate extimation of the shape of the curve, better than measurements with a real steady state source. This way we can measure an accurate reverberation time according to the ISO 3382 standard.

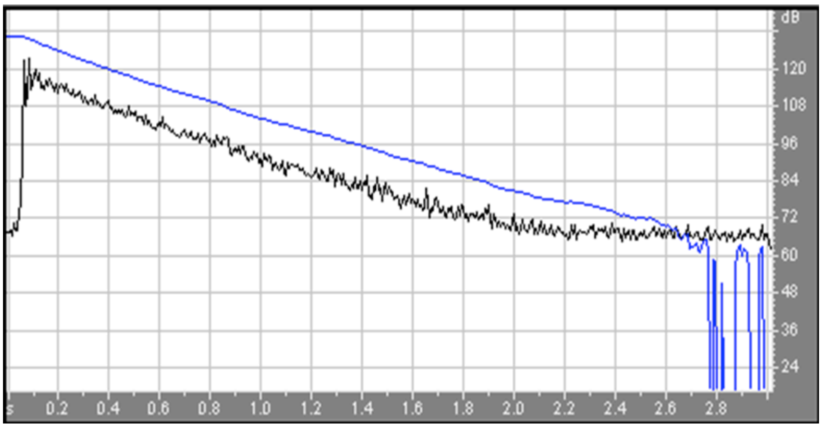


Fig. 4 – Stationary Sound Decay in dB (blue) obtained applying Schroeder BW to an Energetic Impulse Response in dB (black).

Reverberation Time

ISO 3382 doesn't ask to measure reverberation time over a 60 db sound decay because of the signal to noise ratio, we can never wait for a 60 db decay. So the standard tells us to measure **T_{20}**, defined as three times the time required for a 20 db decay. Every reverbation time is referred to a 60 db decay.

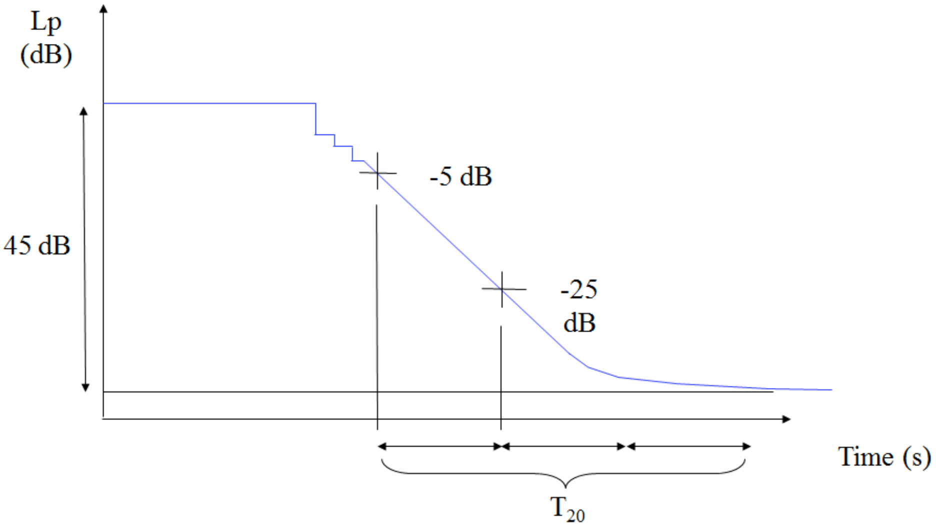


Fig. 5 – ISO 3382 Standard's T_{20} definition.

Actually the standard defines four reverberation times:

- Early Decay Time (**EDT**), extrapolated from 0 to -10 dB (discrete reflections region, not in the reverberant tale)
- Reverberation Time **T_{10}**, extrapolated from -5 to -15 dB
- Reverberation Time **T_{20}**, extrapolated from -5 to -25 dB
- Reverberation Time **T_{30}**, extrapolated from -5 to -35 dB

Early – Late Parameters

The **Early-Late Energy** concept is based on the subdivision of the reverberant tale in two parts: **Useful Energy** and **Detrimental Energy**. This distinction is not correct, our ears love reverb, (if we cut the last part of the response of a room it would sound dry, dead, so this means last part is important too), but we can search a proper balance between these two parts of the response.

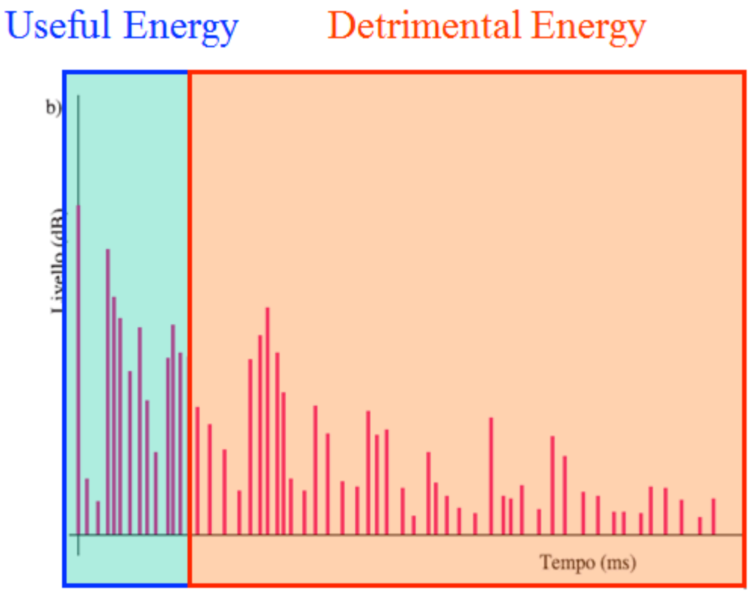


Fig. 6 – Distinction between Useful and Detrimental Energy.

In a room, in spite of the same reverberation time, we can find different responses moving the microphone in different positions, so we need others parameters above reverberation time.

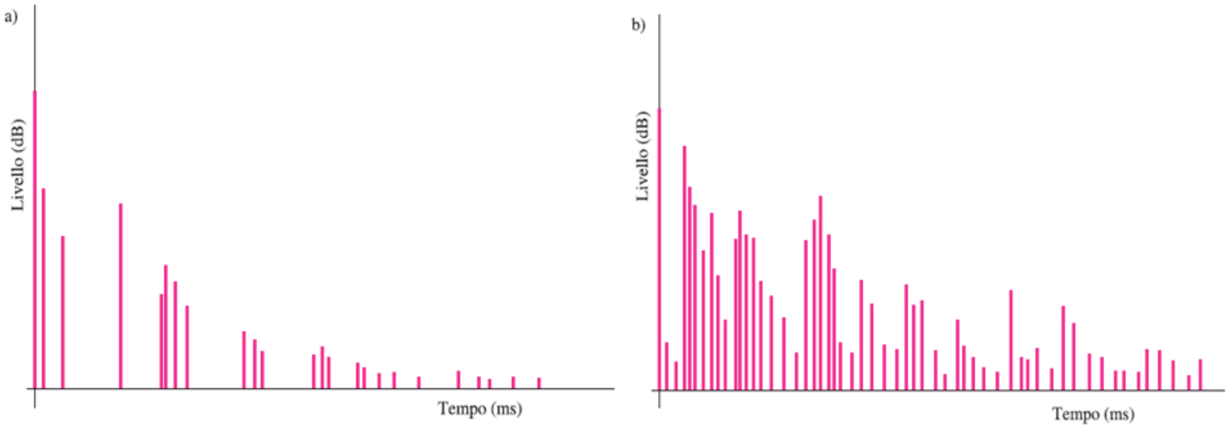


Fig. 7 – IR measured near the source (a) and far from the source (b) in the same room.

- ***Clarity Index*** $\boldsymbol{C_{80}}$ ***(Symphonic Music):***

$$C_{80} = 10 \cdot \lg \left[\frac{\int_{0}^{80ms} p^2(\tau) \cdot d\tau}{\int_{80ms}^{\infty} p^2(\tau) \cdot d\tau} \right] \qquad \text{Optimal value = +/-1 dB} \qquad (4)$$

The boundary between useful and detrimental energy is set to be equal to 80 ms, the integration time of our brain when listening to music.

This is a ratio between early and late energy, and its optimal value is 0 db: when early energy equate late energy.

- ***Clarity Index C_{50} (Speech):***

$$C_{50} = 10 \cdot \lg \left[\frac{\int_0^{50ms} p^2(\tau) \cdot d\tau}{\int_{50ms}^{\infty} p^2(\tau) \cdot d\tau} \right]$$

Optimal value = +/-1 dB (5)

The boundary is set to be equal to 50 ms, the integration time for speech.

- ***Definition Index D:***

$$D = \frac{\int_0^{50ms} p^2(\tau) \cdot d\tau}{\int_0^{\infty} p^2(\tau) \cdot d\tau} \cdot 100 \qquad (6)$$

Equivalent of Clarity, used in North Europe. We use Clarity.

- ***Center Time t_s:***

$$t_s = \frac{\int_0^{\infty} \tau \cdot p^2(\tau) \cdot d\tau}{\int_0^{\infty} p^2(\tau) \cdot d\tau} \qquad (7)$$

There's no pre-defined boundary, it's the point of balance between the energy which is before and the energy which after. Optimal center time to listening to music is 80 ms, for speech is 50 ms.

This parameter can be very important in calibration of rooms for specific kind of music or speech: different genres of music need different balance between early and late energy for optimal listening (ex. Gregorian – Rap music). Same concept for different languages.

Other Parameters

- ***Strenght G:***

$$G = SPL - L_w + 31 \qquad dB \qquad (8)$$

Sound strength is the sound pressure level scaled on the sound power level of our sound source, it is the boost of the room.
This is very important for speech or for singing when the room must provide support. We subtract the SPL at 10 m from our SPL. This means that in the range of 10 m we have a boots because of the reflections.

- ***Inter Aural Cross Correlation IACC:***

$$\rho(\tau) = \frac{\lim_{T\to\infty}\left(\frac{1}{2T}\int_{-T}^{T} p_d(t)\cdot p_s(t+\tau)\,dt\right)}{\lim_{T\to\infty}\left(\frac{1}{2T}\sqrt{\int_{-T}^{T} p_d^2(t)\,dt\cdot\int_{-T}^{T} p_s^2(t+\tau)\,dt}\right)} \qquad (9)$$

IACC is a spatial parameter defined as the correlation between the signal measured at the entrance of the ears of a dummy-head or of a true human head. Lower **IACC** means a high perception difference between the two microphones and so a good stereophony and envelopment.

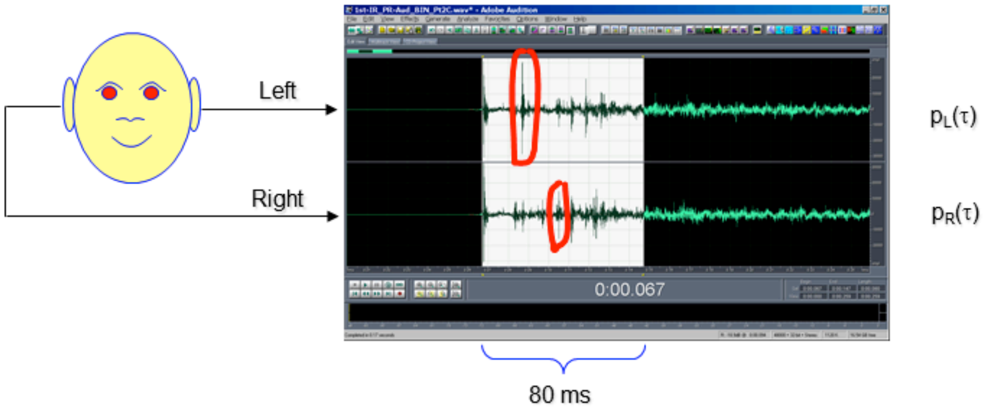


Fig. 8 – Binaural IR measurements. Calibrated on 80 ms for music.

Some reflections arrive louder to only one of the two microphones: those are lateral reflections and they help to perceive spatiality. Some others are the same in both the IR and can't help to recognize direction of the sound: those are ceiling reflections for example.

- ***Lateral Fraction J_{LF}:***

$$J_{LF} = \frac{\int_{5ms}^{80ms} h_Y^{\,2}(\tau)\cdot d\tau}{\int_{0ms}^{80ms} h_W^{\,2}(\tau)\cdot d\tau} \qquad (10)$$

Lateral Fraction is a spatial parameter defined as the ratio between the sound coming from a side and the whole sound. For measurement we can use a pressure-velocity microphone.

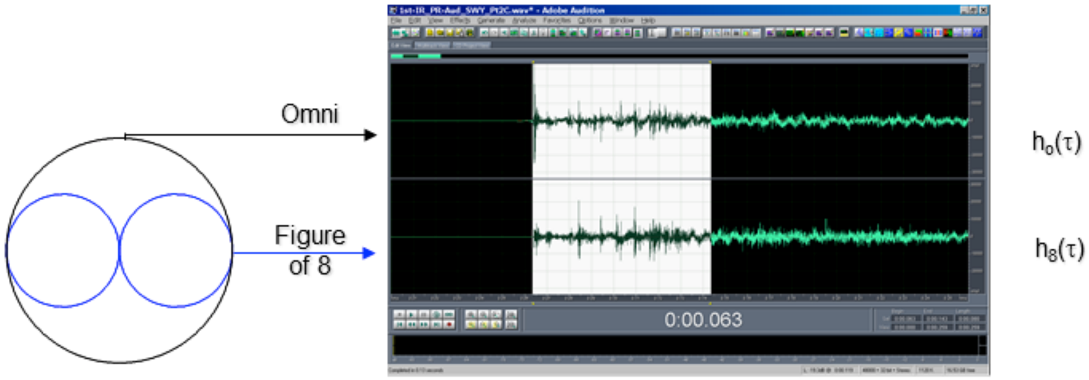


Fig. 9 – Pressure-velocity microphone measurements.

This time the two IR channels are not Left-Right but Omni-Figure of eight. The integral start after 5 ms so, even if we haven't placed the microphone facing the source (from top in Fig. 11), direct sound doesn't affect our measurements.

Usually only one of the two spatial parameters is measured:

$$J_{FL} \approx 1 - IACC \qquad (11)$$

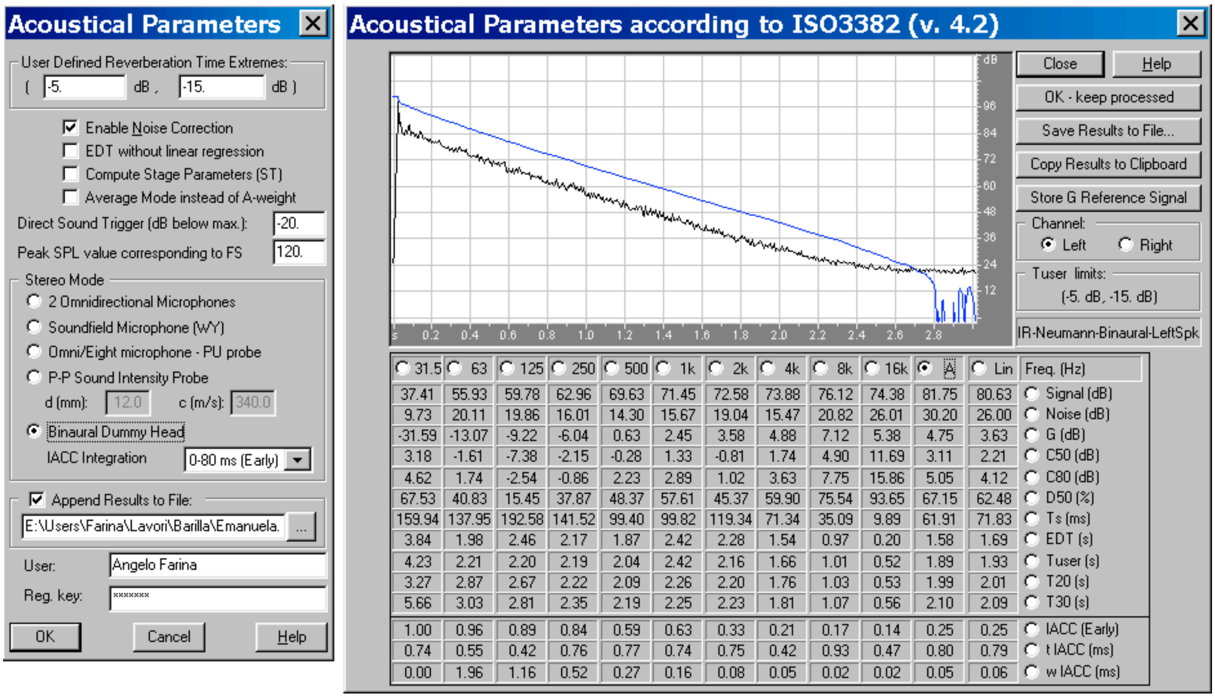
Measurements reproducibility

Testing different Dummy-Heads (for IACC) and Pressure-velocity microphones (for Lateral Fraction) in the same room with the same source and comparing the results we can see that these measurements are not very scientific and each dummy head/microphone can give a different result.

This means we can trust this measurements for spatial analysis only when used comparatively, we can't take them as absolute.

The Aurora Acoustical Parameters and STI Plugins

Aurora software has a nice plugin for IR's analysis which can give us all the parameters contained in the ISO 3382 standard. We can also isolate single frequency and observe IR and parameters resulting from the analysis.



User Defined Reverberation Time Extremes: (-5. dB , -15. dB)

☑ Enable Noise Correction
☐ EDT without linear regression
☐ Compute Stage Parameters (ST)
☐ Average Mode instead of A-weight

Direct Sound Trigger (dB below max.): -20.
Peak SPL value corresponding to FS: 120.

Stereo Mode
○ 2 Omnidirectional Microphones
○ Soundfield Microphone (WY)
○ Omni/Eight microphone - PU probe
○ P-P Sound Intensity Probe
d (mm): 12.0 c (m/s): 340.0
● Binaural Dummy Head
IACC Integration: 0-80 ms (Early)

☑ Append Results to File: E:\Users\Farina\Lavori\Barilla\Emanuela.

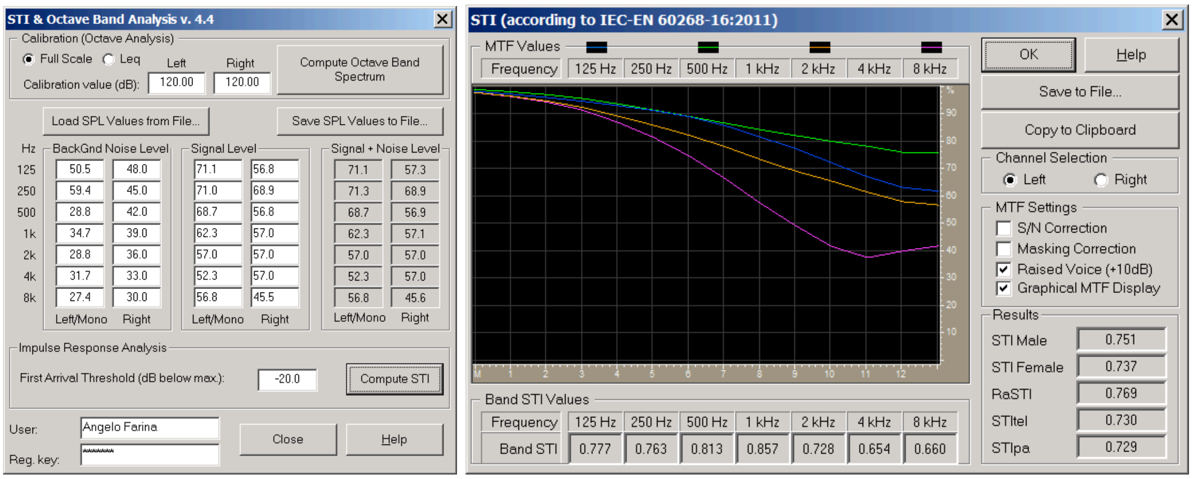
User: Angelo Farina
Reg. key: *******

Channel: ● Left ○ Right
Tuser limits: (-5. dB, -15. dB)
IR-Neumann-Binaural-LeftSpk

○ 31.5	○ 63	○ 125	○ 250	○ 500	○ 1k	○ 2k	○ 4k	○ 8k	○ 16k	● A	○ Lin	Freq. (Hz)
37.41	55.93	59.78	62.96	69.63	71.45	72.58	73.88	76.12	74.38	81.75	80.63	○ Signal (dB)
9.73	20.11	19.86	16.01	14.30	15.67	19.04	15.47	20.82	26.01	30.20	26.00	○ Noise (dB)
-31.59	-13.07	-9.22	-6.04	0.63	2.45	3.58	4.88	7.12	5.38	4.75	3.63	○ G (dB)
3.18	-1.61	-7.38	-2.15	-0.28	1.33	-0.81	1.74	4.90	11.69	3.11	2.21	○ C50 (dB)
4.62	1.74	-2.54	-0.86	2.23	2.89	1.02	3.63	7.75	15.86	5.05	4.12	○ C80 (dB)
67.53	40.83	15.45	37.87	48.37	57.61	45.37	59.90	75.54	93.65	67.15	62.48	○ D50 (%)
159.94	137.95	192.58	141.52	99.40	99.82	119.34	71.34	35.09	9.89	61.91	71.83	○ Ts (ms)
3.84	1.98	2.46	2.17	1.87	2.42	2.28	1.54	0.97	0.20	1.58	1.69	○ EDT (s)
4.23	2.21	2.20	2.19	2.04	2.42	2.16	1.66	1.01	0.52	1.89	1.93	○ Tuser (s)
3.27	2.87	2.67	2.22	2.09	2.26	2.20	1.76	1.03	0.53	1.99	2.01	○ T20 (s)
5.66	3.03	2.81	2.35	2.19	2.25	2.23	1.81	1.07	0.56	2.10	2.09	○ T30 (s)
1.00	0.96	0.89	0.84	0.59	0.63	0.33	0.21	0.17	0.14	0.25	0.25	○ IACC (Early)
0.74	0.55	0.42	0.76	0.77	0.74	0.75	0.42	0.93	0.47	0.80	0.79	○ t IACC (ms)
0.00	1.96	1.16	0.52	0.27	0.16	0.08	0.05	0.02	0.02	0.05	0.06	○ w IACC (ms)

Fig. 10 – Acoustical Parameters plugin.

Another very useful Aurora plugin is STI, which computes the Speech Transmission Index:



Calibration (Octave Analysis): ● Full Scale ○ Leq

Calibration value (dB): Left 120.00 Right 120.00

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	50.5	48.0	71.1	56.8	71.1	57.3
250	59.4	45.0	71.0	68.9	71.3	68.9
500	28.8	42.0	68.7	56.8	68.7	56.9
1k	34.7	39.0	62.3	57.0	62.3	57.1
2k	28.8	36.0	57.0	57.0	57.0	57.0
4k	31.7	33.0	52.3	57.0	52.3	57.0
8k	27.4	30.0	56.8	45.5	56.8	45.6

Impulse Response Analysis
First Arrival Threshold (dB below max.): -20.0

User: Angelo Farina
Reg. key: ********

MTF Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz

Band STI Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Band STI	0.777	0.763	0.813	0.857	0.728	0.654	0.660

Channel Selection: ● Left ○ Right

MTF Settings
☐ S/N Correction
☐ Masking Correction
☑ Raised Voice (+10dB)
☑ Graphical MTF Display

Results

STI Male	0.751
STI Female	0.737
RaSTI	0.769
STItel	0.730
STIpa	0.729

Fig. 11 – Acoustical Parameters plugin.

Speech Transmission Index

1 - The STI method

STI is the acronym of "Speech Transmission Index", it's defined as an estimate of speech intelligibility, and it's standardized in IEC 60268/16.

Intelligibility means capacity of a receiver to listen correctly phrases and words pronounced by a source.

The index does not define directly intelligibility, but a low value defines a loss of information necessary to understanding speech correctly.

Intelligibility is a fundamental factor to evaluate the quality of communication inside a room or through telephonic equipment.

Many factors that influence intelligibility do exist:

- If the speaker's voice comes through an electro-acoustic system, there are factors that can affect intelligibility (for examples frequency response and distortion of the system).
- The acoustic characteristics of the environment are the reverberation (or reflections), the presence of background noise, the echoes, etc.

We will study the two most significant factors: background noise and reverberation.

The STI method is based on the MTF concept, which defines how much the modulation of a carrier signal (one-octave-band-filtered noise) is reduced when such a test signal passes through the system under test.

The MTF factor is defined as the ratio of the carrier's modulation at the receiver (ex. 50%) and the carrier's modulation at the source (ex. 100%).

The carrier signal is pink noise filtered in one octave band. The measurement is repeated for 7 octave bands (125 Hz to 8 kHz). We call ***f*** the center frequency of the carrier.

The modulation is applied as a periodic variation of the carrier's intensity between 0 and the maximum value (hence, initially, the modulation is at 100%). The modulation frequency is called ***F***, and ranges between 0.63 Hz and 12.5 Hz.

Hence, we can measure a large number of ***MTF(f,F)*** values, given by all the possible combinations of ***f*** and ***F***.

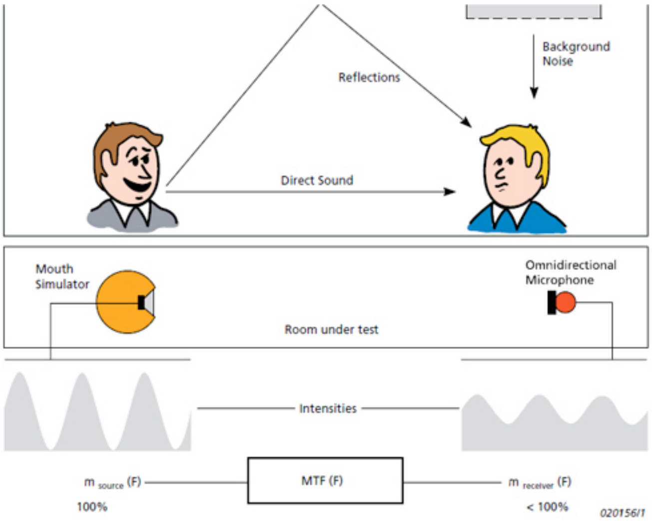


Fig.1 – The initial modulation of the carrier is reduced by the propagation and by the environmental noise.

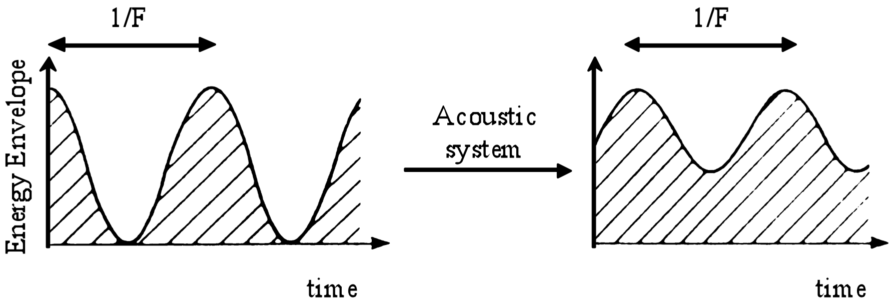


Fig.2 – Sound propagation through an acoustic system reduces the carrier's modulation.

Once the MTF value is found at every value of ***f*** and ***F***, the values referring to each octave band are first averaged. Then, a weighted average of these "Band STI" values is performed, employing averaging factors depending on the gender of the talker (male or female).

The resulting "total" STI is defined as a number bounded between zero and one.

The maximum value is one (100%) and defines perfect intelligibility, the minimum value is zero (0%) and means that the modulation is not audible.

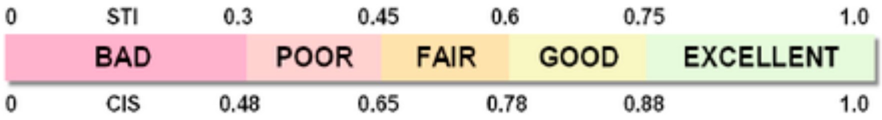


Fig.3 – STI and CIS scale of values .

Another reference scale called CIS (Common Intelligibility Scale) exists, based on a mathematical relation with STI:

$$CIS = 1 + \log(STI) \qquad (1)$$

2 - MTF from impulse Response

The IEC standard defines a method to measure STI values.

The test signal has spectral characteristics and directivity similar to the human voice.

f defines the octave bands (7 bands between 125Hz and 8kHz), while ***F*** defines the modulation frequency, where the human mouth opens and closes (between 0.63 Hz and 12.5 Hz).

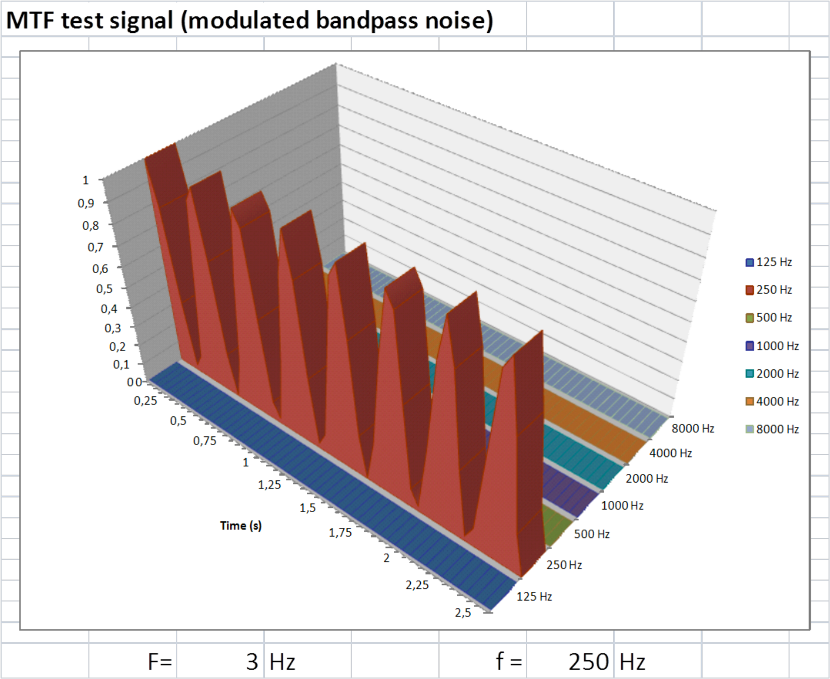


Fig.4 – MTF test signal.

The MTF matrix can be computed by a single impulse response (IR).

MTF(f,F) is defined by the combination of two factors: reverberation (first factor) and signal/noise ratio (second factor).

$$m(f,F)=\frac{\int_0^\infty h_f^{\,2}(\tau)\cdot\exp(-j\cdot 2\cdot\pi\cdot F\cdot\tau)\cdot d\tau}{\int_0^\infty h_f^{\,2}(\tau)\cdot d\tau}\cdot\frac{1}{1+10^{\left(\frac{L_{noise,f}-L_{signal,f}}{10}\right)}} \qquad (2)$$

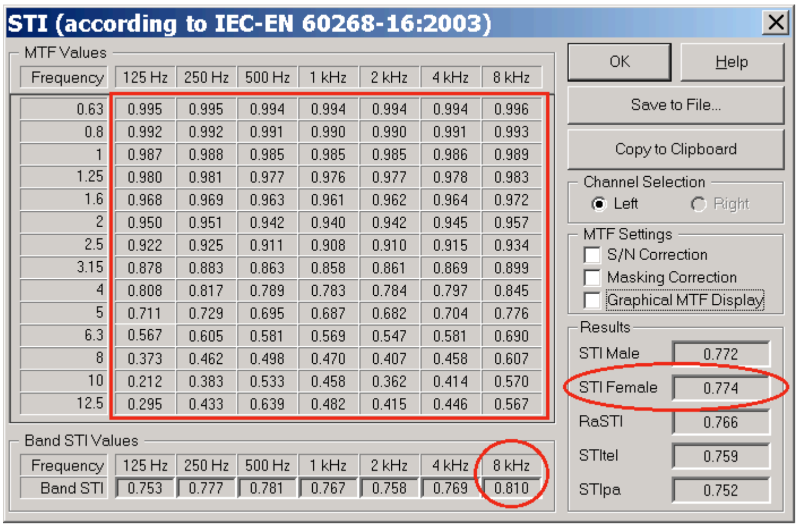
Shroeder's equation computes the first factor, related to reverberation, called ***m'(F)*** (the calculation is based on the octave-band filtered impulse response ***h***$_f$).

The loudness and duration of the reverberant tail decrease the value of STI. However, for a given reverberation time, its effect will be worst at higher modulation frequency F, hence it is usual to see the value of MTF(f,F) to decrease with F.

At high frequency usually the S/N ratio is better than at low frequency, hence typically MTF(f,) increases with f. The second factor of equation (2) does not depend on the modulation frequency F, so, if the low value of MTF is caused by a S/N ratio problem, the MTF curve is NOT decreasing with F, as it happens instead when the reverberation is the limiting factor to intelligibility.

Hence, the experienced acoustician immediately understands, looking at the table or chart of MTF(f,F), if the intelligibility problems are caused by excessive reverberation or excessive noise (or both).

The male voice has more energy at low frequency, the female voice instead has more energy at high frequency, in fact usually the female voice is more audible than the male one, as it profits of better S/N ratio at high frequency.



STI (according to IEC-EN 60268-16:2003)

MTF Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
0.63	0.995	0.995	0.994	0.994	0.994	0.994	0.996
0.8	0.992	0.992	0.991	0.990	0.990	0.991	0.993
1	0.987	0.988	0.985	0.985	0.985	0.986	0.989
1.25	0.980	0.981	0.977	0.976	0.977	0.978	0.983
1.6	0.968	0.969	0.963	0.961	0.962	0.964	0.972
2	0.950	0.951	0.942	0.940	0.942	0.945	0.957
2.5	0.922	0.925	0.911	0.908	0.910	0.915	0.934
3.15	0.878	0.883	0.863	0.858	0.861	0.869	0.899
4	0.808	0.817	0.789	0.783	0.784	0.797	0.845
5	0.711	0.729	0.695	0.687	0.682	0.704	0.776
6.3	0.567	0.605	0.581	0.569	0.547	0.581	0.690
8	0.373	0.462	0.498	0.470	0.407	0.458	0.607
10	0.212	0.383	0.533	0.458	0.362	0.414	0.570
12.5	0.295	0.433	0.639	0.482	0.415	0.446	0.567

Band STI Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Band STI	0.753	0.777	0.781	0.767	0.758	0.769	0.810

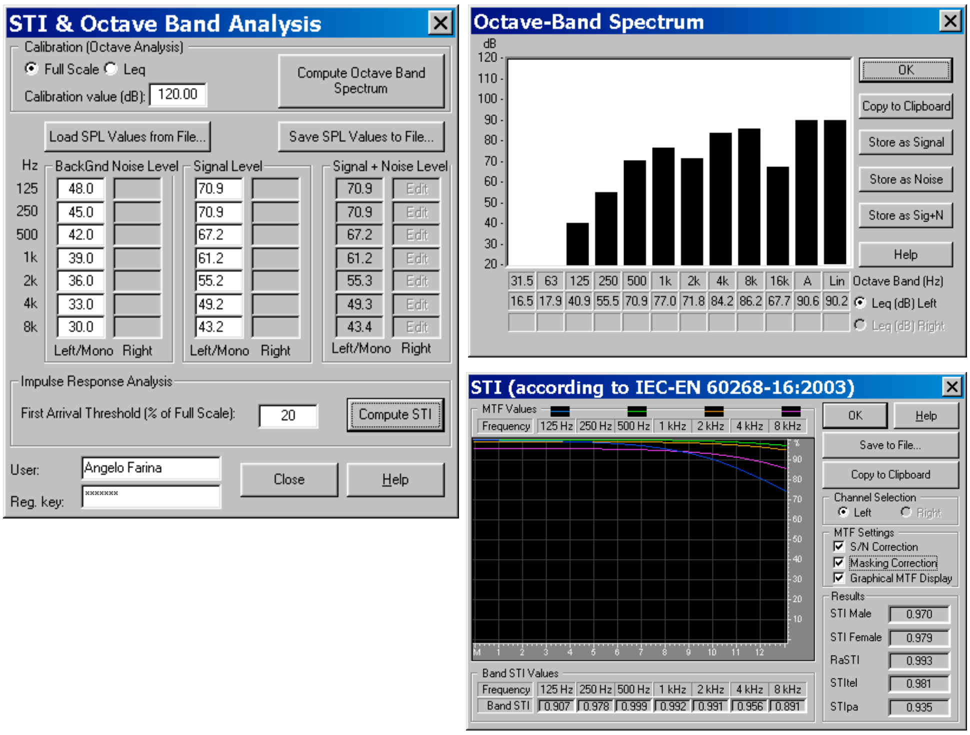
Channel Selection
● Left ○ Right

MTF Settings
☐ S/N Correction
☐ Masking Correction
☐ Graphical MTF Display

Results

STI Male	0.772
STI Female	0.774
RaSTI	0.766
STItel	0.759
STIpa	0.752

Fig.5 – m(F) matrix, m(F) average values for octave bands and total STI values for male and female voices.



STI & Octave Band Analysis

Calibration (Octave Analysis)
● Full Scale ○ Leq
Calibration value (dB): 120.00

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	48.0		70.9		70.9	Edit
250	45.0		70.9		70.9	Edit
500	42.0		67.2		67.2	Edit
1k	39.0		61.2		61.2	Edit
2k	36.0		55.2		55.3	Edit
4k	33.0		49.2		49.3	Edit
8k	30.0		43.2		43.4	Edit

Impulse Response Analysis
First Arrival Threshold (% of Full Scale): 20

User: Angelo Farina
Reg. key: *******

Octave-Band Spectrum

Octave Band (Hz)	31.5	63	125	250	500	1k	2k	4k	8k	16k	A	Lin
● Leq (dB) Left	16.5	17.9	40.9	55.5	70.9	77.0	71.8	84.2	86.2	67.7	90.6	90.2
○ Leq (dB) Right												

STI (according to IEC-EN 60268-16:2003)

MTF Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz

Channel Selection
● Left ○ Right

MTF Settings
☑ S/N Correction
☑ Masking Correction
☑ Graphical MTF Display

Band STI Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Band STI	0.907	0.978	0.999	0.992	0.991	0.956	0.891

Results

STI Male	0.970
STI Female	0.979
RaSTI	0.993
STItel	0.981
STIpa	0.935

Fig.6 – Post processing of impulse response.

STI requires to measure MTF(f,f9 on seven octave bands and 14 modulation frequencies, RaSTI (Rapid STI) is measured at just two octave bands (500Hz and 2kHz) and with a smaller number of modulation frequencies.

Nowadays the technology is able to compute in the same time the complete MTF matrix, and the full STI – so the use of RaSTI is now obsolete, such as other "rapid" versions of ST, named STIpa and STItel – it is always preferable to measure the full STI, as the measurement time is indeed always the same..

3 – Transducers: mouth simulator

A mouth simulator has characteristics in level and directivity similar to those of a real human talker

.

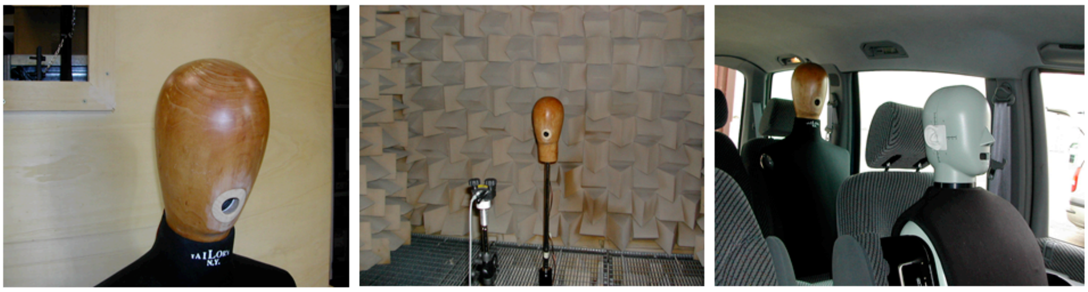


Fig.7 – Simulators build inside a wooden head (employing low-cost parts) and in (expensive) plastic head.

The validity of the mouth simulator is confirmed by a means of anechoic directivity tests.

Mouth simulator's spectrum can be adjusted to become perfectly equal to the standardized human voice spectrum, but the directivity is always slightly different.

The spectrum of emitted test signal should correspond to the ITU T-P50 standard.

The overall SPL should be 60 dB(A) at 1m, on axis, for measurements compliant with IEC 60268-16 standard.

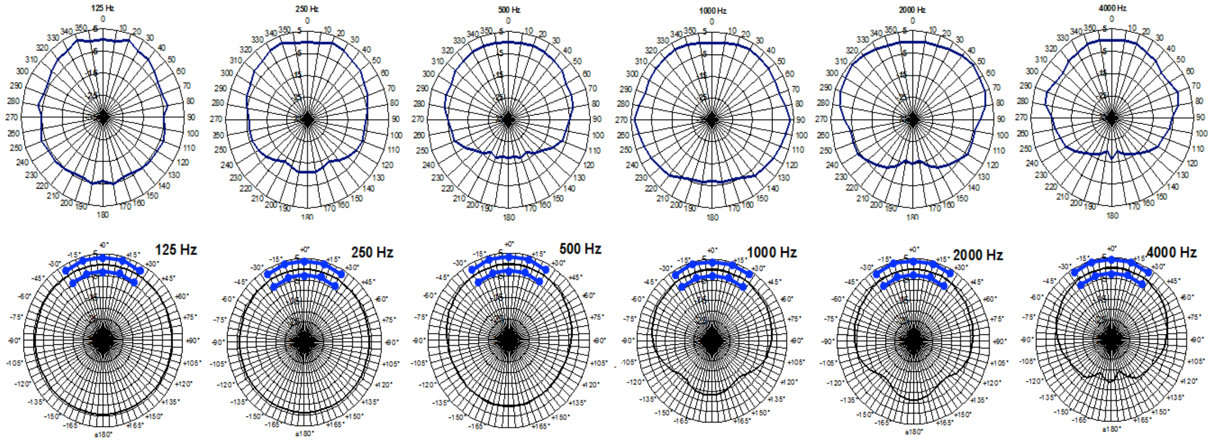


Fig.9 – directivity (up real, down simulator, blue line ITU limits).

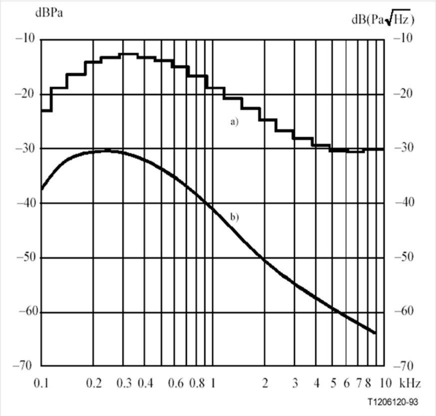
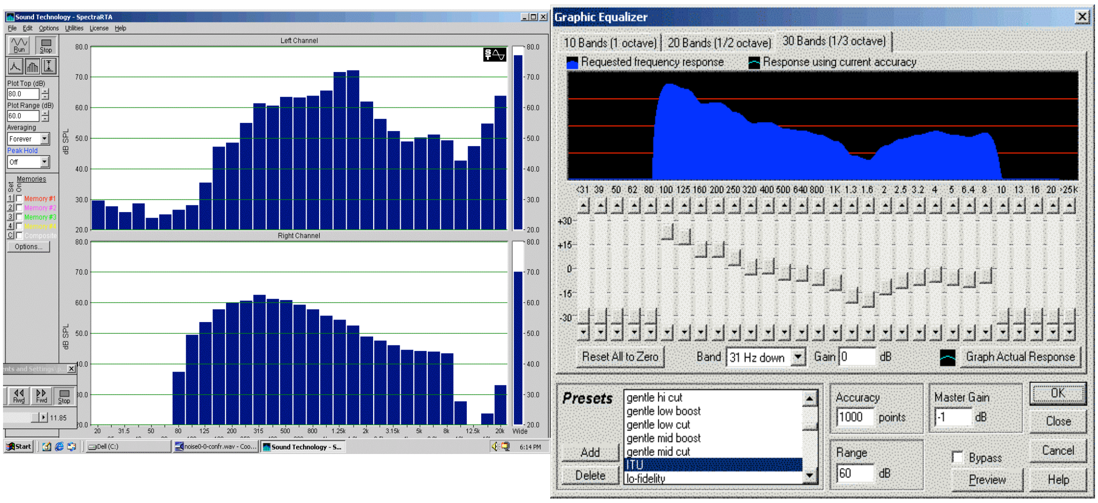


Fig.10 – target spectrum according to ITU P50.

The equalization of the simulator is easily operated by means of the graphic equalizer included in Adobe Audition.

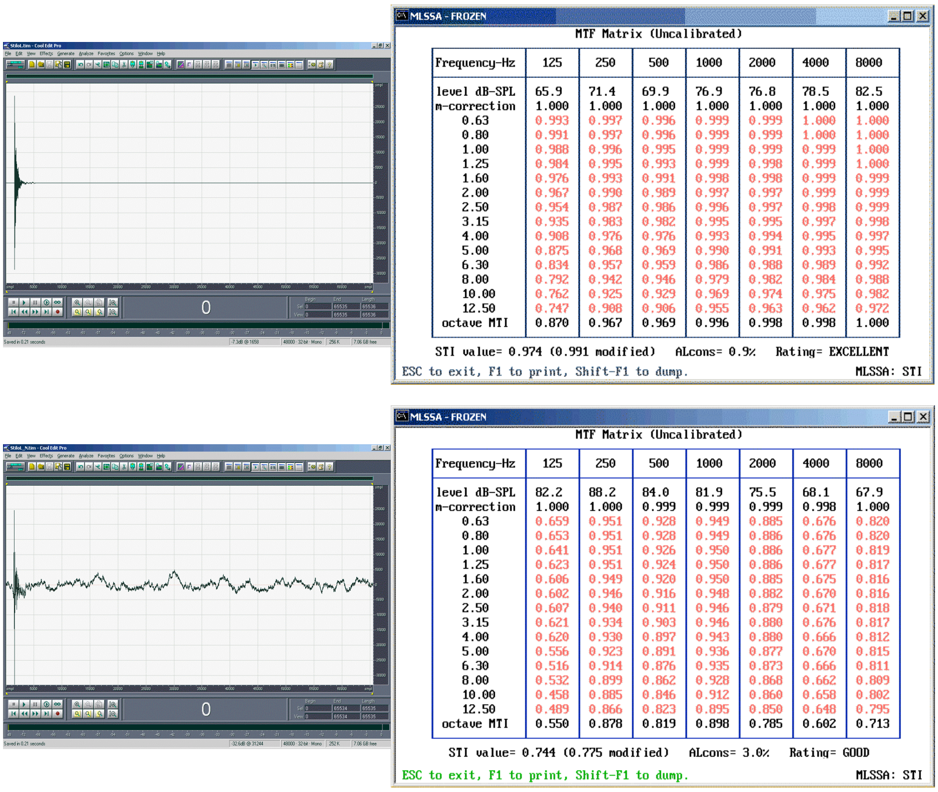


Band 31 Hz down
Gain 0 dB
Accuracy 1000 points
Master Gain -1 dB
Range 60 dB

Fig.11 – Graphic equalizer.

The test signal is prefiltered, so that the frequency response measured at 1m in front of the mouth, complies with the IEC spectrum (or, better, with the ITU P50 standard, which specifies values in 1/3 octave bands).

The measured IR is saved as a WAV file.



MTF Matrix (Uncalibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB-SPL	65.9	71.4	69.9	76.9	76.8	78.5	82.5
m-correction	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.63	0.993	0.997	0.996	0.999	0.999	1.000	1.000
0.80	0.991	0.997	0.996	0.999	0.999	1.000	1.000
1.00	0.988	0.996	0.995	0.999	0.999	0.999	1.000
1.25	0.984	0.995	0.993	0.999	0.998	0.999	1.000
1.60	0.976	0.993	0.991	0.998	0.998	0.999	0.999
2.00	0.967	0.990	0.989	0.997	0.997	0.999	0.999
2.50	0.954	0.987	0.986	0.996	0.997	0.998	0.999
3.15	0.935	0.983	0.982	0.995	0.995	0.997	0.998
4.00	0.908	0.976	0.976	0.993	0.994	0.995	0.997
5.00	0.875	0.968	0.969	0.990	0.991	0.993	0.995
6.30	0.834	0.957	0.959	0.986	0.988	0.989	0.992
8.00	0.792	0.942	0.946	0.979	0.982	0.984	0.988
10.00	0.762	0.925	0.929	0.969	0.974	0.975	0.982
12.50	0.747	0.908	0.906	0.955	0.963	0.962	0.972
octave MTI	0.870	0.967	0.969	0.996	0.998	0.998	1.000

```
STI value= 0.974 (0.991 modified)   ALcons= 0.9%   Rating= EXCELLENT
ESC to exit, F1 to print, Shift-F1 to dump.                  MLSSA: STI
```

MTF Matrix (Uncalibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB-SPL	82.2	88.2	84.0	81.9	75.5	68.1	67.9
m-correction	1.000	1.000	0.999	0.999	0.999	0.998	1.000
0.63	0.659	0.951	0.928	0.949	0.885	0.676	0.820
0.80	0.653	0.951	0.928	0.949	0.886	0.676	0.820
1.00	0.641	0.951	0.926	0.950	0.886	0.677	0.819
1.25	0.623	0.951	0.924	0.950	0.886	0.677	0.817
1.60	0.606	0.949	0.920	0.950	0.885	0.675	0.816
2.00	0.602	0.946	0.916	0.948	0.882	0.670	0.816
2.50	0.607	0.940	0.911	0.946	0.879	0.671	0.818
3.15	0.621	0.934	0.903	0.946	0.880	0.676	0.817
4.00	0.620	0.930	0.897	0.943	0.880	0.666	0.812
5.00	0.556	0.923	0.891	0.936	0.877	0.670	0.815
6.30	0.516	0.914	0.876	0.935	0.873	0.666	0.811
8.00	0.532	0.899	0.862	0.928	0.868	0.662	0.809
10.00	0.458	0.885	0.846	0.912	0.860	0.658	0.802
12.50	0.489	0.866	0.823	0.895	0.850	0.648	0.795
octave MTI	0.550	0.878	0.819	0.898	0.785	0.602	0.713

```
STI value= 0.744 (0.775 modified)   ALcons= 3.0%   Rating= GOOD
ESC to exit, F1 to print, Shift-F1 to dump.                  MLSSA: STI
```

Fig.12 – MLSSA calculates MTF for “no noise” (up) and “noise” (down).

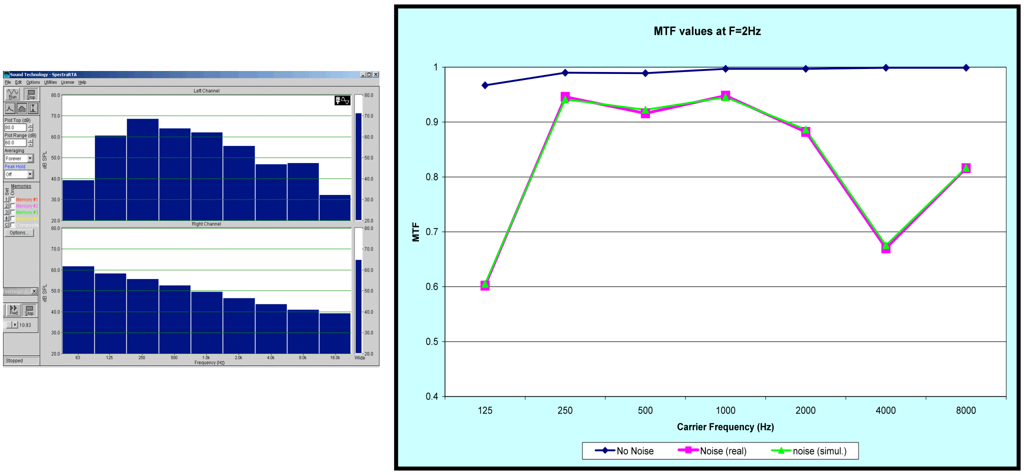


Fig.13 – Equation (2) has high accuracy because differences of real and simulated MTF are minimal.

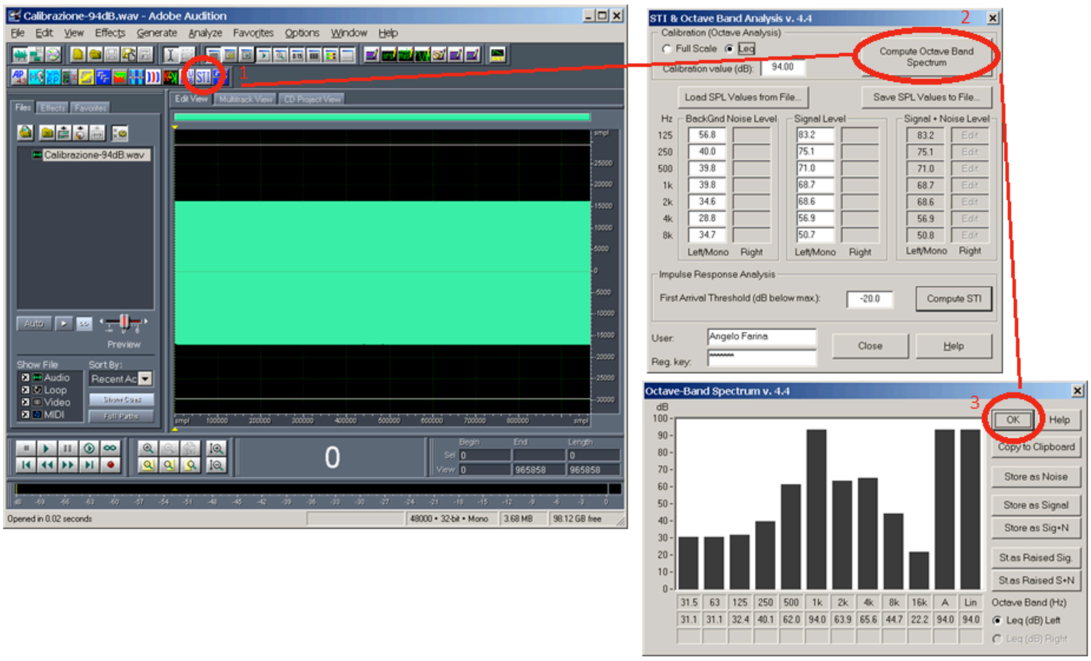
.

4 – Noise-free IR Method (Aurora Plugins)

“Aurora Plugins Suite” can be used for a complete STI measurement, requiring four steps, as follows:

4.1 - Calibration

The calibrator is fitted over the microphone and a 1-minute recording of the 1kHz test signal at 94 dB is done.



STI & Octave Band Analysis v. 4.4

Calibration (Octave Analysis): ○ Full Scale ● Leq

Calibration value (dB): 94.00

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	56.8		83.2		83.2	Edit
250	40.0		75.1		75.1	Edit
500	39.8		71.0		71.0	Edit
1k	39.8		68.7		68.7	Edit
2k	34.6		68.6		68.6	Edit
4k	28.8		56.9		56.9	Edit
8k	34.7		50.7		50.8	Edit

Impulse Response Analysis

First Arrival Threshold (dB below max): -20.0

User: Angelo Farina

Reg. key: *******

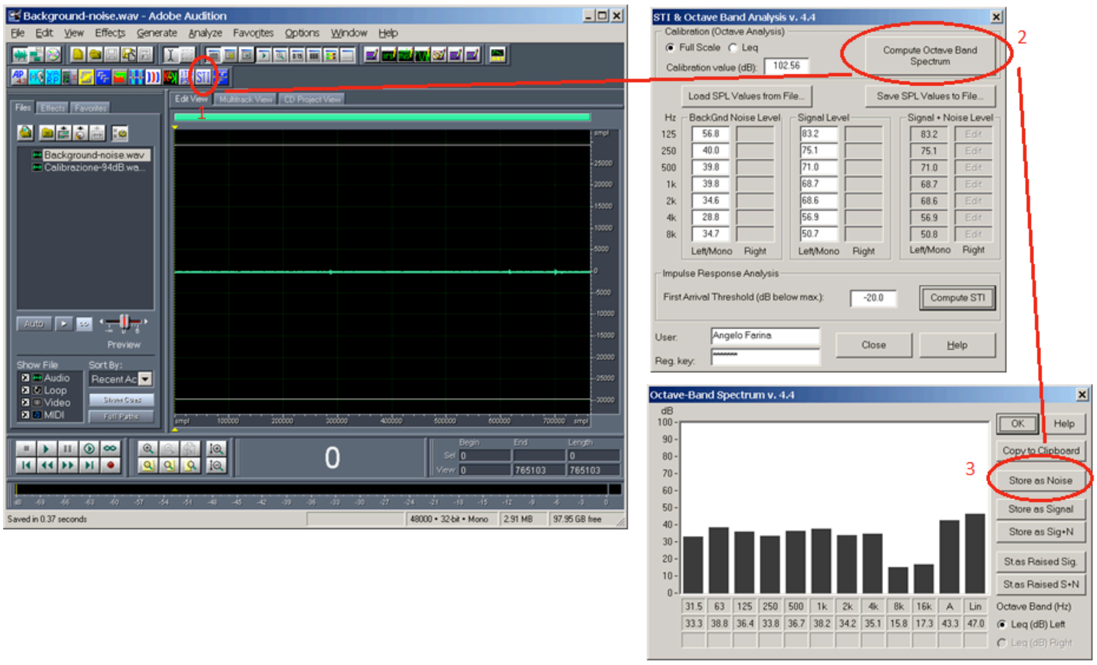
Octave-Band Spectrum v. 4.4

Octave Band (Hz)	31.5	63	125	250	500	1k	2k	4k	8k	16k	A	Lin
● Leq (dB) Left	31.1	31.1	32.4	40.1	62.0	94.0	63.9	65.6	44.7	22.2	94.0	94.0
○ Leq (dB) Right												

Fig.14 –microphone calibration.

The STI plugin is invoked, forcing the Leq value to be 94 dB at 1 kHz. Later on, the Full Scale value will be left untouched, so that the microphone is now calibrated in absolute SPL.

4.2 - Background Noise spectrum



Calibration (Octave Analysis): ● Full Scale ○ Leq

Calibration value (dB): 102.56

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	56.8		83.2		83.2	Edit
250	40.0		75.1		75.1	Edit
500	39.8		71.0		71.0	Edit
1k	39.8		68.7		68.7	Edit
2k	34.6		68.6		68.6	Edit
4k	28.8		56.9		56.9	Edit
8k	34.7		50.7		50.8	Edit

Impulse Response Analysis

First Arrival Threshold (dB below max): -20.0

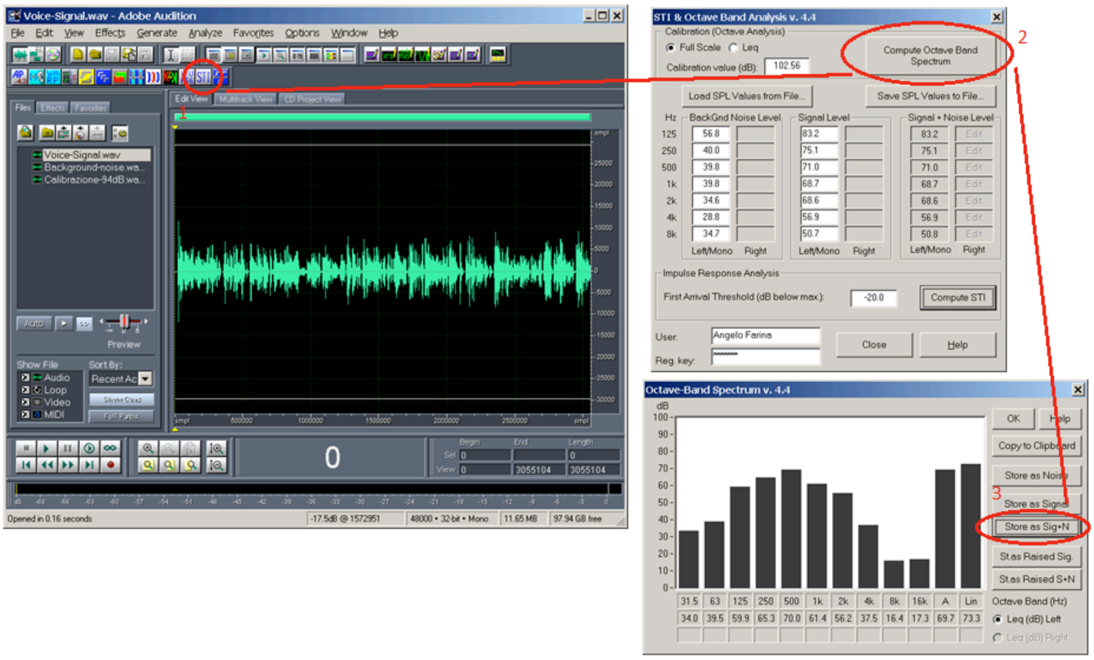
User: Angelo Farina

Reg. key: *******

Octave Band (Hz)	31.5	63	125	250	500	1k	2k	4k	8k	16k	A	Lin
● Leq (dB) Left	33.3	38.8	36.4	33.8	36.7	38.2	34.2	35.1	15.8	17.3	43.3	47.0
○ Leq (dB) Right												

Fig.15 – Record background noise and store it as noise.

4.3 - Signal spectrum



Calibration (Octave Analysis): ● Full Scale ○ Leq

Calibration value (dB): 102.56

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	56.8		83.2		83.2	Edit
250	40.0		75.1		75.1	Edit
500	39.8		71.0		71.0	Edit
1k	39.8		68.7		68.7	Edit
2k	34.6		68.6		68.6	Edit
4k	28.8		56.9		56.9	Edit
8k	34.7		50.7		50.8	Edit

Impulse Response Analysis

First Arrival Threshold (dB below max): -20.0

User: Angelo Farina

Reg. key: *******

Octave Band (Hz)	31.5	63	125	250	500	1k	2k	4k	8k	16k	A	Lin
● Leq (dB) Left	34.0	39.5	59.9	65.3	70.0	61.4	56.2	37.5	16.4	17.3	69.7	73.3
○ Leq (dB) Right												

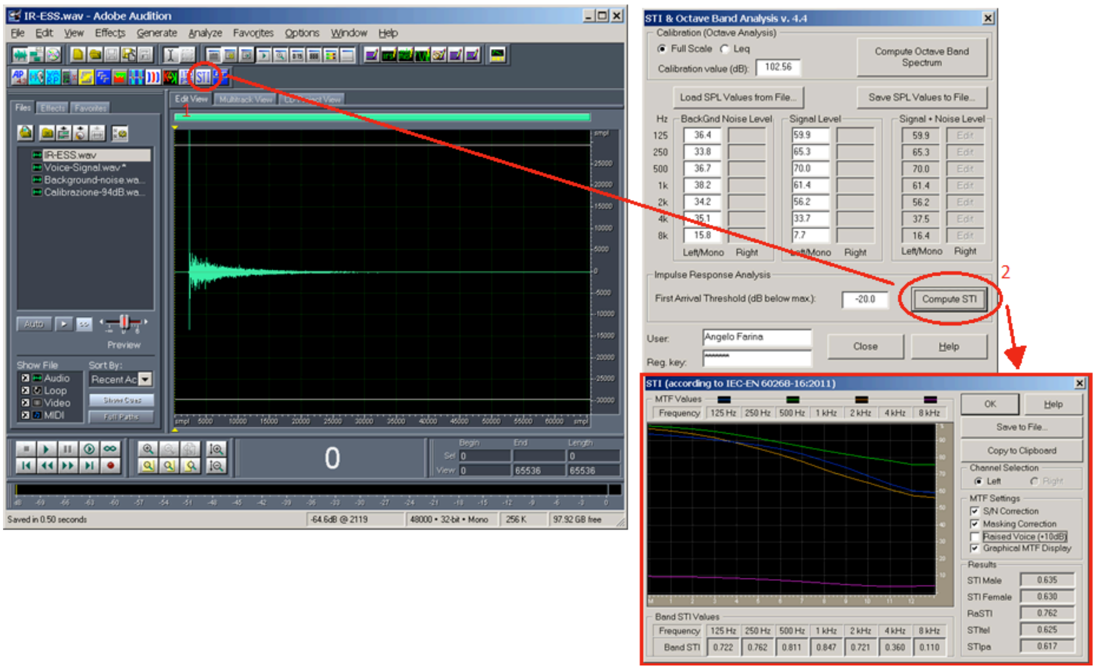
Fig.16 – Record the pre-equalized wide-band noise (voice spectrum calibrated at 1m) and store it as signal plus noise.

4.4 - Impulse Response (noise-free, ESS method)

Now, a standard I measurement is performed, generating the ESS signal, playing it through the mouth simulator, recording the system response at the microphone, and finally performing deconvolution (convolution with the inverse sweep). The resulting IR is stored in a WAV file, usually normalized to full scale (as the absolute SPL of a noise-free IR is meaningless)

From this IR, the Aurora STI plugin computes the MTF values for every frequency, also taking optionally into account the S/N ratio.

The MTF values are also weighting averaged, to compute male and female STI (the weightings are different).



STI & Octave Band Analysis v. 4.4

Calibration (Octave Analysis): ● Full Scale ○ Leq

Calibration value (dB): 102.56

Hz	BackGnd Noise Level Left/Mono	BackGnd Noise Level Right	Signal Level Left/Mono	Signal Level Right	Signal + Noise Level Left/Mono	Signal + Noise Level Right
125	36.4		59.9		59.9	Edit
250	33.8		65.3		65.3	Edit
500	36.7		70.0		70.0	Edit
1k	38.2		61.4		61.4	Edit
2k	34.2		56.2		56.2	Edit
4k	35.1		33.7		37.5	Edit
8k	15.8		7.7		16.4	Edit

Impulse Response Analysis – First Arrival Threshold (dB below max.): -20.0

User: Angelo Farina

Reg. key: *******

STI (according to IEC-EN 60268-16:2011)

MTF Settings:
☑ S/N Correction
☑ Masking Correction
☐ Raised Voice (+10dB)
☑ Graphical MTF Display

Channel Selection: ● Left ○ Right

Band STI Values

Frequency	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Band STI	0.722	0.762	0.811	0.847	0.721	0.360	0.110

Results

STI Male	0.635
STI Female	0.630
RaSTI	0.762
STItel	0.625
STIpa	0.617

Fig.17 – Aurora processes the IR, “Noise” and “Signal” to compute MTF averaged coefficients and male and female STI. The plugin also computes RaSTI, STItel and STIpa

The nice thing of this approach is the possibility to evaluate “what happens if” – for example, if the voice is raised by +10dB (a specific check box is available for this).

Or, by processing the IR, it is possible to simulate a room treatment which reduces the reverberation time, or to suppress a single discrete echo.