1.1 STI evaluation with Monkey Forest

Monkey Forest is able to perform STI calculations from impulse responses according to the international standard IEC 60268-16, third edition from 2003. The calculations involve various corrections to account for level dependent masking from one octave band to another, the absolute hearing threshold and a redundancy correction for adjacent octaves. The final result are two revised STI values, one for female and one for male voices. For the sake of compatibility and comparability to older versions and other programs, all corrections can be disabled separately. The original STI version with slightly different octave weighting factors and without redundancy correction is also available.

The following explanations assume that the reader is familiar with the concept of modulation transfer functions and the procedure to obtain the STI.

1.1.1 Impulse response and SNR

The STI is basically dependent on two properties of a room: The impulse response and the prevalent back ground noise relative to the speaker level, each one considered in the seven octave bands from 125 Hz to 8 kHz. While it is theoretically possible to derive the signal to noise ratio from the impulse response measurement (provided that the excitation signal's spectral distribution and level match speech), this approach would be rather unpractical. The goal of an impulse response measurement normally is to achieve an SNR as high as possible. Preferably, the SNR should also be independent of frequency to allow for trouble-free evaluation of other room-acoustical parameters, most notably the reverberation time. Monkey Forest features extensive tools for giving the excitation signal any desired spectral shape. These are normally used to equalize the loudspeaker and to compensate the background noise. Giving the excitation signal the shape of human voice would lead to lacking SNR in some frequency regions, especially the 63 Hz octave, normally included in room-acoustical measurements.

For this reason, the impulse response and the SNR per octave are handled separately by Monkey Forest. The impulse responses can be measured using all possibilities to increase the SNR (emphasis, extended length, synchronous averages, volume) without having to worry about matching spectral distribution of voice and playback level.

The SNR values have to be determined in two separate measurements (taking advantage, for example, of the SPL menu in the spectral domain) with and without speech.

The separate handling of impulse response and SNR also bears the advantage of allowing to determine how strong the background noise worsens the octave-wise and overall results by enabling and disabling the inclusion of the SNR in the calculations.

1.1.2 Conditions for correct results

Several conditions must be met to achieve truthful and reliable STI results with Monkey Forest and to allow smooth operation:

- The sampling rate must be at least 32 kHz to safely include the 8 kHz band. If the sampling rate of your IR is lower and you are merely interested in the RASTI, perform an `Alt Edit/Here comes more/Sample rate conversion` to bring the sample rate up.

- The recovered impulse response should have at least a length of 1.6 seconds, even if the expected reverberation time is much shorter. If this requirement is violated, errors will be introduced in the m-values of the lower modulation frequencies. The 1.6 seconds correspond to a full period of 0.63 Hz, the lowest modulation frequency involved in the STI calculation. If the impulse response is shorter than that, the Fourier transform for this particular frequency cannot be performed over the full period, resulting in m-values considerably above 0 even if the input is pure noise (which correspondents to infinitely bad acoustics).

- However, the impulse response should not be extended with zeros to achieve a certain length (for example, the 1.6 s limit). This is not only unnecessary because it will yield the same result as with the original IR, but it will dramatically increase the calculation time. The reason for this is that in the zero-padded region, denormalized numbers with extremely small values circulate in the IIR-filter stages. This is very painful for the FPU (floating point unit), storing and retrieving those numbers can slow down operation up to 8 times. When the original IR has a length of less than 1.6 seconds, it is better to leave it this way and accept the warning issued by Monkey Forest when calling the STI function.

- The recovered impulse response should have an SNR of at least 20 dB in all octave bands, preferably much higher than that. If this requirement is not fulfilled, the prevalent background noise not only enters into the final result via the SNR table, but also a second time via the impulse response, leading to worse STI results than are reality.

- The impulse response must be either broadband or suitably pre-filtered into octave bands. Suitably pre-filtered means that the octave-wide bandpass filters applied from

125 Hz to 8 kHz should adhere to the IEC 1260 (appropriate filters are automatically selected when using the `Alt Edit / Filter` menu). If the range of the resulting bands is larger than necessary for the STI calculation (e.g. including the 63 Hz octave), Monkey Forest will automatically pick up the right bands.

1.1.3 The Menu STI Brimborium

The STI function is available in the time domain and in the impulse response domain. It is invoked by typing Alt Info / t. The first thing that pops up on screen is the welcome box which is divided into two sections:

1.1.3.1 Preparation

`# Setup` leads you to the usual file select menu where you can store or recall STI setups including all switches and entries accessible in all menus related to the STI, including the S/N-Table:

`Table SN` This will guide you to a submenu where you can view and modify the octave-wise signal- and noise levels. They are important if any of three corrections later described are active.

Fig. 1: STI welcome box.

`Fast processing` Monkey Forest can process the input impulse responses in two ways:

The standard is by filtering them into octave bands, squaring them and performing discrete Fourier transforms at the exact STI modulation frequencies. All this is done maintaining the full original sampling rate. If the impulse responses are already filtered into octaves, the first step will of course be omitted. The fast processing option is not available in this case. If, however, the input impulse responses are still broadband, the first step of the STI calculation is to automatically apply low-ripple Chebychev octave bandpass filters of 8th order, replacing the original time signal with the seven octaves per input channel. This first filtering step is much the same as if it had been activated by using the multiple-filter bank in *Alt Edit / Filter* and leads to identical results.

The *fast processing* works a bit different. It won't change the broadband impulse responses on screen, but instead submits them to a subsequent filtering and downsampling from higher to lower octaves, which greatly relieves the computational burden and yields almost identical results in much shorter time.

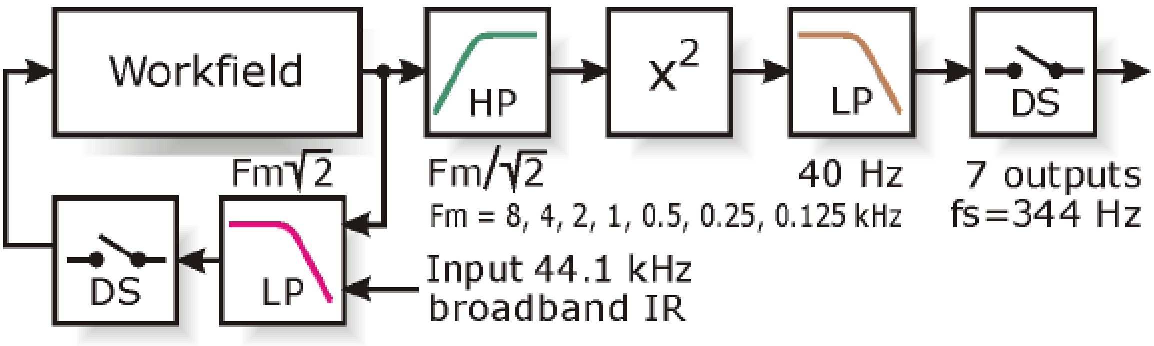


Fig. 2: Filtering and downsampling scheme in „fast processing" mode, with input 44.1 kHz sampling rate as example.

Each broadband impulse response is first low-pass filtered with the upper edge of the 8 kHz octave band and stored in a temporal field. If the original sampling rate is equal or higher than 64 kHz, than a suitable downsampling is applied. From here, the following two steps are repeated seven times:

1) The contents of the temporal field is passed through a high-pass filter with a corner frequency equaling the lower edge of the octave band, squared, low pass filtered (Butterworth 4th order, f_{CUT} = 40 Hz) and finally sampled down to the final sampling rate. The result (sampling rate for 44.1 kHz input: 344 Hz) is stored in a relatively small array which will later be processed by discrete Fourier transforms (DFTs) for the STI calculation.

2) The contents of the temporal field is now low-pass-filtered with the upper edge of the adjacent smaller octave and then downsampled by a factor of 2. The result is again stored in the same field.

This fast processing yields the STI in much shorter time for two reasons:

1) The filtering is completed in less time.

2) The time to perform the DFTs shrinks to 1/128 (and even less due to handling inside the processor caches) compared to the standard case. As a matter of fact, it is now so fast that it has a negligible contribution to the overall processing time.

As additional advantage, the fast processing option leaves the impulse responses on screen unaltered and may also be applied to a single channel in a multi-channel impulse responses. To do so, activate the desired channel (by directly typing its number or using the `*/` keys) and switch `all channels` off before calling the STI. When activating the STI in the `impulse response` domain, MF always uses fast processing.

There is no reason why you shouldn't work with the fast processing, unless you want to try out special filters (for example, FIR filtering with higher selectivity) or compare the results as a reference with those computed by other software.

There is, however, a small difference to standard processing with full sampling rate because the filter shape of the octave bandpass filters changes slightly when decreasing the sampling rate. This becomes most evident for the lowest band, the 125 Hz octave:

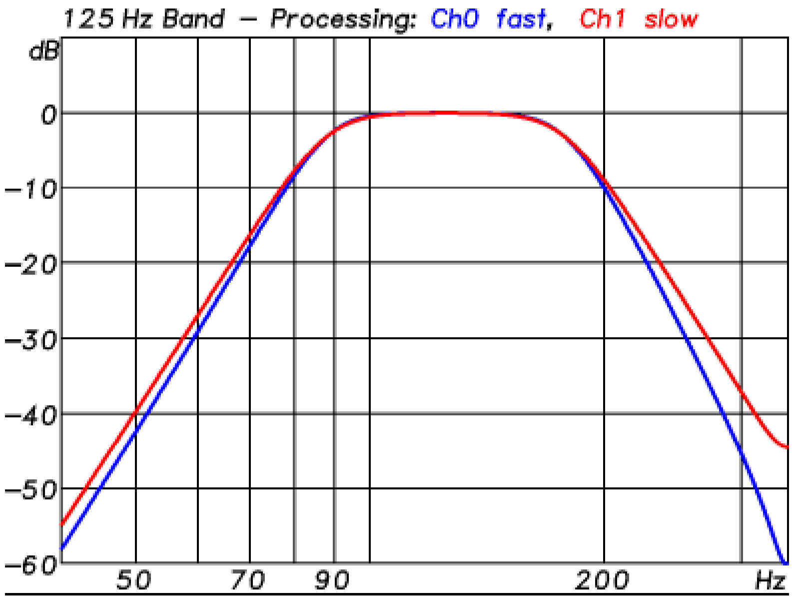


Fig. 3: Filter curves for the 8th order Butterworth highpass/lowpass combination in fast (blue) and standard processing mode.

The selectivity of the filter operating at reduced sampling rate is slightly higher than of the one operating at the original sampling rate. This can cause minor differences in the calculated modulation transfer function, should high level components exist in the adjacent bands. The passband is also slightly divergent for both versions. However, both filters are adequate for the separation of the octave bands and the final STI results normally do not differ by more than 0.001.

1.1.3.2 Display parameters

Yes, after the extensively presented fast processing feature: we're still in the initial STI welcome box. Its second part is dedicated to the display parameters which can be individually switched on and off for the result window. The activated STI parameters will be shown in one row, while the Rating, %Alcons and CIS, if activated, will be shown in separate rows, each one derived from each activated STI parameter.

The only two values which really matter are the STIr male and female. These are the two parameters calculated by following the guidelines in the 2003 version of the IEC 60268-16 standard. They are activated as default. The other STI values have only been included to allow for comparisons with older measurements and/or results of other programs.

The STIPA, which only considers 2 instead of the normal 14 modulation frequencies per octave band, is not even intended to be derived from impulse responses, but rather from simultaneous analysis of a continuously running special excitation noise. However, it might be interesting to compare the results of genuine STIPA measuring gear with the results based on impulse responses and separate S/N tables. Apart from the fact that only two modulation frequencies per octave band enter into the result, the STIPA is calculated exactly as the STIr, using the same weighting factors and redundancy corrections.

The "old STI" is calculated according to the original definition of the STI dating back to the sixties and seventies. There is no redundancy correction for neighboring octave bands and the weighting factors for the octave-wise modulation transfer indices (MTI) are slightly different than for the new revised STIr. However, all other corrections (SNR, masking and hearing threshold) are applied in the same way as for the STIr. This also holds for the STIPA and the RASTI. In order to compare results of older measurements, all these corrections should thus be switched off.

The RASTI (rapid STI) is an obsolescent condensed version of the STI. It only considers four modulation frequencies in the 500 Hz band and five in the 2 kHz band. It has been introduced a long time ago when computers still were slow and the full STI calculation could take some minutes. Because only two octave bands are involved, its calculation really would be faster than the full STI. However, Monkey Forest does not even consider the case "only RASTI" selected, meaning that the filtering is always done for all seven octaves. As a matter of fact, including RASTI in the display parameters even bears a small speed penalty, because the modulation frequencies for the 2 kHz band are outside of the standard STI scheme and thus the m-contributions for the RASTI in this band have to be calculated by five separate DFTs. This speed penalty, though small, should be enough to discourage the user to bother with the RASTI at all. If even the predictability of speech intelligibility using the full STI has already been cast in doubt, its castrated companion RASTI must be definitely worse as intelligibility predictor.

The STITEL is not yet implemented and might continue so in the future, unless someone convinces the program author of its usefulness.

1.1.3.3 Go for it!

Finally, after carefully choosing the desired parameters, pressing the Enter key starts the STI evaluation, provided that the IR has a compatible sampling rate and is either unfiltered or filtered in octave bands, spanning at least the range from 125 Hz to 8 kHz.

1.1.4 The Signal / Noise Table

The signal/noise table has to contain valid values if any of the level dependent corrections prescribed by the latest STI standard are activated. The values can be entered manually, or, if a calibrated microphone is connected to the measurement hardware, by performing an SPL measurement on site (`Alt A /SPL` in the spectral domain) with octave-wide adding (`smooth -> Parameter -> Bandwidth : 1/1 octave)` and activating the `copy to STI S/N` option in the last row. After each measurement, the octave levels will be stored in the signal- or in the noise row, depending on your choice.

Signal/noise levels

Mode : sig/sig+noise

Band	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Signal	78 dB	83 dB	87 dB	86 dB	84 dB	79 dB	68 dB
Noise	75 dB	74 dB	70 dB	69 dB	71 dB	65 dB	55 dB

Quit

Fig. 4: Signal/noise table in mode „sig“.

Both the signal and the noise measurement should of course be performed under the expected operating conditions. For the signal measurement, two different cases are handled:

1) The signal level is measured beforehand with the absence of noise. This holds for many measurements, where acquiring impulse responses and the speech levels is performed with still no audience present. In this case, the `mode` switch must be set to `sig`.

2) The signal level is measured under true operating conditions, this means the audience is present, has already taken some drinks and become rather boisterous. This means that the SPL measurement of the speech signal in reality captures the sum of speech and noise. In this case, the `mode` switch in the menu should accordingly be set to

`sig+noise`. This will evoke a third line between the Sig+N and Noise rows in which the signal levels calculated according to

$$S = 10 \cdot \log(10^{S+N/10} - 10^{N/10})$$

are shown. These are the values which will enter into the STI calculation in this mode.

Signal/noise levels

Mode : sig/sig+noise

Band	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Sig+N	78 dB	83 dB	87 dB	86 dB	84 dB	79 dB	68 dB
Signal	75 dB	82.4 dB	86.9 dB	85.9 dB	83.8 dB	78.8 dB	67.8 dB
Noise	75 dB	74 dB	70 dB	69 dB	71 dB	65 dB	55 dB

Quit

Fig. 5: Signal / Noise table in mode sig+noise.

Make sure that the signal+noise levels are always above the noise levels. If they become almost the same, Monkey forest will clamp the derived signal-only level to the value of the noise level minus 20 dB.

1.1.5 The Result Menu

After pressing <Enter> in the welcome box, the impulse responses (up to 4) are processed and the MTF matrices are filled. The resulting STI values and their associated ratings, %ALcons and CIS values, if activated, are presented in the result window. Here, the various corrections related to the octave wise signal and noise levels can be switched on and off to watch their influence on the final results. The inclusion of SNR and level-dependent masking between adjacent octave bands of course worsens the result compared to the noiseless case. In contrast, the hearing threshold correction has practically no influence unless the signal levels are extremely low.

The signal/noise table is still accessible as a sub menu and any modifications in it will be directly reflected in the results when returning to the result window.

The rating for each STI value will be:

STI value	rating
0.75 - 1	excelent
0.6 - 0.75	good
0.45 – 0.6	fair
0.3 - 0.45	poor
0.15 – 0.3	bad
0 - 0.15	ugly

Tab 1: STI ratings.

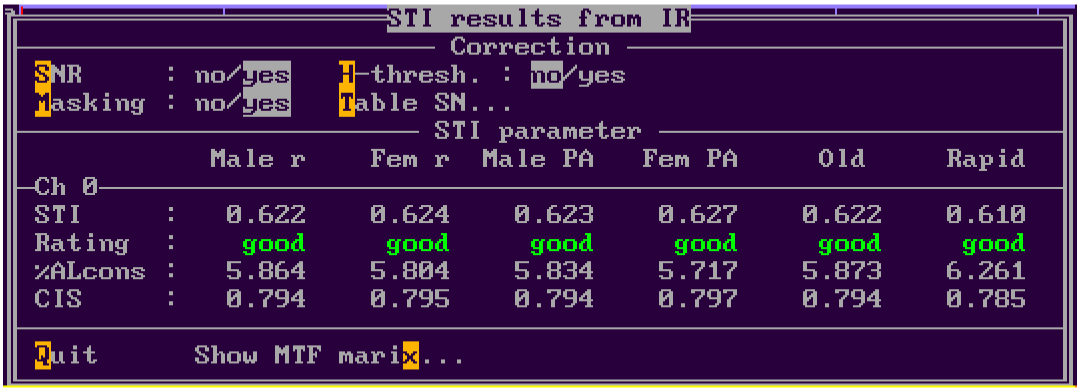


Fig. 6: STI result window.

The percentage of articulation loss for consonants (%ALcons) does not involve a separate calculation process, but is merely derived from the individual STI values using the following empirical equation according to Farrel Becker:

$$\%AlCons = 170.5405 \cdot e^{-5.419 \cdot STI}, \; 0 <= \%AlCons <= 100$$

The common intelligibility scale (CIS) is even simpler to calculate:

$$CIS = 1 + \log(STI), \; 0 <= CIS <= 1$$

1.1.6 The MTF Matrix window and the STI calculation in detail

The MTF matrix window is accessible from within the result windows and presents all m-values for one channel that enter into the final result for the STI value specified in the `display mode.` The display mode RASTI is only available when RASTI has been chosen in the STI welcome box as one of the desired results. If not, the special DFTs for the 2 kHz band have not been calculated and are thus not available.

Again, as in the result window, all corrections can be individually switched on and off and their effect immediately observed in the row `m-reduction`. There are three different corrections, presented now in detail.

The most important one is to take into account the SNR. When the correct signal and noise values have been evaluated and are present in the signal & noise level table, this correction should always be active, unless the *noiseless STI* shall be calculated to explicitly exclude the influence of the noise. The octave-specific SNR correction applied to all m values within one band is:

$$m_{CORSNR} = m \cdot \frac{1}{1+10^{\frac{-S/N}{10}}}$$

The second correction takes masking into account. It will only have a significant influence if the speaker level in one band is considerably higher than in the next band (which can occur if the PA is badly equalized). To simplify things, only the lower band *k-1* is taken into account. Its masking on the next band *k* is modeled as a steady slope with constant dB/octave value:

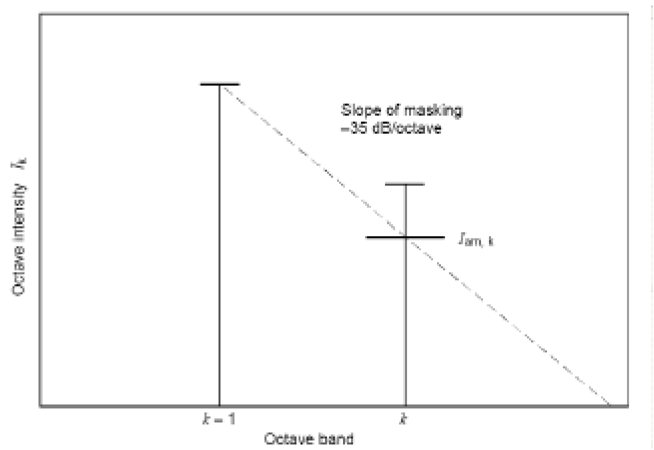


Fig. 7: Auditory masking from band k-1 to adjacent band k.

Unfortunately, the slope of this value depends on the octave level in the masker band:

Octave level dB	46-55	56-65	66-75	76-85	86-95	>95
Slope of masking	−40	−35	−25	−20	−15	−10
Auditory masking factor	0,000100	0,000316	0,003162	0,010000	0,031622	0,100000

Tab 2: Level specific slope depending on the masking level and corresponding auditory masking factor *amf*.

With the help of the masker level S_{k-1} in band *k-1* and the corresponding factor *amf* taken from Tab 2, the influence on the masked band *k* is modeled as interfering noise intensity:

$$I_{am,k} = 10^{\frac{S_{k-1}}{10}} \cdot amf$$

The third correction takes into account that the SNR would have to increase to maintain constant intelligibility if the level drops towards the absolute hearing threshold. This is modeled as a fixed octave-band specific interfering intensity calculated from the *absolute reception threshold* levels given in Tab 3:

$$I_{rs,k} = 10^{\frac{L_{rs,k}}{10}}$$

The influence of this term in the overall correction will be small if the octave-band signal levels are high. With the help of I_k, which is the intensity of the speech signal in band k,

$$I_k = 10^{\frac{S_k}{10}}$$

the combined m-reduction $mcor_k$ of a particular octave band k, taking into account all three corrections, becomes:

$$mcor_k = \frac{1}{1 + 10^{\frac{-S/N}{10}}} \cdot \frac{I_k}{I_k + I_{am,k} + I_{rs,k}}$$

This is the octave-wise value shown in the `m-reduction` row. All 14 m-values of the particular band k are multiplied with this correction:

$$m'_{k,f} = m_{k,f} \cdot mcor_k$$

After these modifications, the STI is calculated. To do so, first all corrected m-values are transformed to apparent signal-to-noise ratios:

$$SNR_{k,f} = 10 \cdot \log \frac{m'_{k,f}}{1 - m'_{k,f}} dB$$

These SNR values are clipped to the range ±15 dB and then transformed into transmission indices (TI):

$$TI_{k,f} = \frac{SNR_{k,r} + 15dB}{30dB}$$

The modulation transfer index MTI for each octave is then calculated by averaging all TI values for one octave band:

$$MTI_k = \frac{1}{14} \sum_{f=1}^{14} TI_{k,f}$$

From these octave-wise MTI_k, the final STI is now calculated by multiplying the MTI_k with a weighting factor and summing the products up:

$$STI = \sum_{n=1}^{7} a_n \cdot MTI_n - \sum_{n=1}^{6} b_n \sqrt{MTI_n \cdot MTI_{n+1}}$$

The essential part is the first sum. The second sum is a redundancy correction, considering that a good result in one band can partially compensate a worse result in the next band. The factors a and ß are the octave-band specific weighing factors, since the 3rd edition of the 60268-16 there are different ones for males and females:

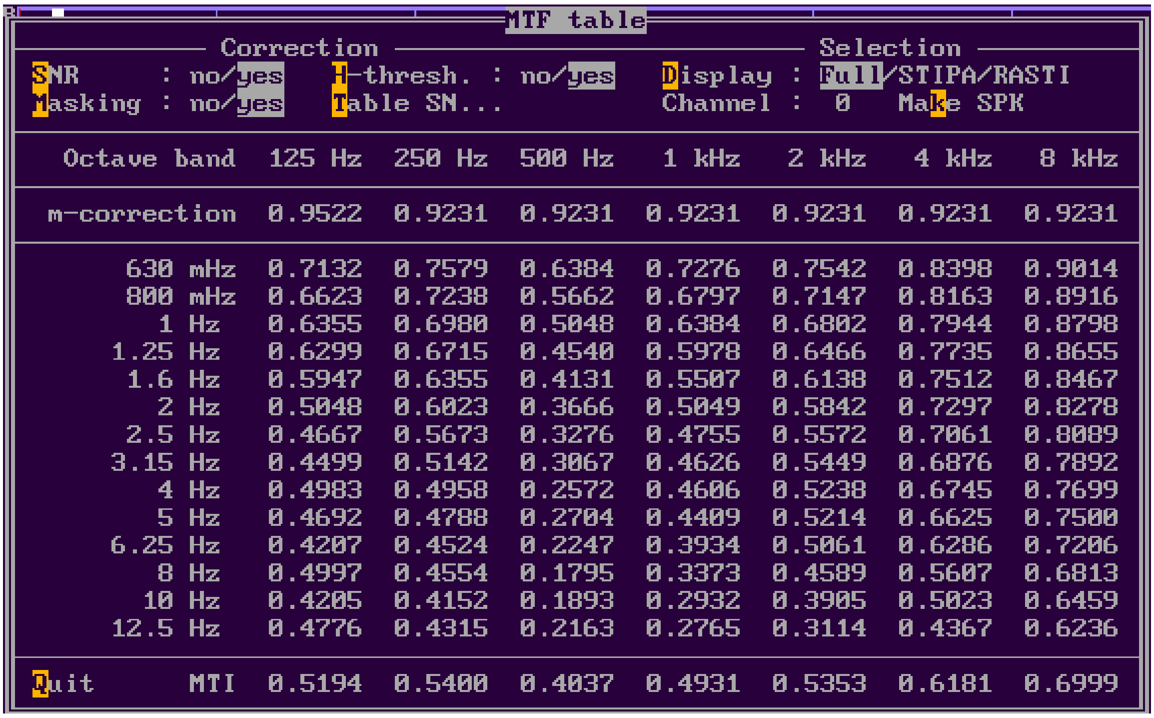
Octave band Hz		125	250	500	1 k	2 k	4 k	8 k
Males	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–
Females	α	–	0,117	0,223	0,216	0,328	0,250	0,194
	β	–	0,099	0,066	0,062	0,025	0,076	–
Absolute reception threshold dB	$L_{rs,k}$	46	27	12	6,5	7,5	8	12

Tab 3: Octave-band specific weighing factors for males and females and levels for the absolute reception threshold correction.

If the STI analysis has been carried out for more than one channel, the displayed channel is selectable.

For post-processing in Excel or other software that accepts tabular ASCII input, the window contents, as any other text window in Monkey Forest, can be appended to the file MF.PRO (the name of this file can be modified in the menu *`Alt Utility /More files /Protocol file`*). Along with the window contents, the signal's headline will be copied to this ASCII file in order to facilitate identifying the chunks of information written into the file. Use an ASCII editor to cut out the desired pieces of information.

The MTFs can also be transformed into a 7 band spectrum, press the k key (for *`Make SPK`*) to do so. The spectrum will then be available upon entering into the spectral domain.



MTF table

Correction: SNR : no/yes, H-thresh. : no/yes, Masking : no/yes, Table SN...

Selection: Display : Full/STIPA/RASTI, Channel : 0, Make SPK

Octave band	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
m-correction	0.9522	0.9231	0.9231	0.9231	0.9231	0.9231	0.9231
630 mHz	0.7132	0.7579	0.6384	0.7276	0.7542	0.8398	0.9014
800 mHz	0.6623	0.7238	0.5662	0.6797	0.7147	0.8163	0.8916
1 Hz	0.6355	0.6980	0.5048	0.6384	0.6802	0.7944	0.8798
1.25 Hz	0.6299	0.6715	0.4540	0.5978	0.6466	0.7735	0.8655
1.6 Hz	0.5947	0.6355	0.4131	0.5507	0.6138	0.7512	0.8467
2 Hz	0.5048	0.6023	0.3666	0.5049	0.5842	0.7297	0.8278
2.5 Hz	0.4667	0.5673	0.3276	0.4755	0.5572	0.7061	0.8089
3.15 Hz	0.4499	0.5142	0.3067	0.4626	0.5449	0.6876	0.7892
4 Hz	0.4983	0.4958	0.2572	0.4606	0.5238	0.6745	0.7699
5 Hz	0.4692	0.4788	0.2704	0.4409	0.5214	0.6625	0.7500
6.25 Hz	0.4207	0.4524	0.2247	0.3934	0.5061	0.6286	0.7206
8 Hz	0.4997	0.4554	0.1795	0.3373	0.4589	0.5607	0.6813
10 Hz	0.4205	0.4152	0.1893	0.2932	0.3905	0.5023	0.6459
12.5 Hz	0.4776	0.4315	0.2163	0.2765	0.3114	0.4367	0.6236
Quit MTI	0.5194	0.5400	0.4037	0.4931	0.5353	0.6181	0.6999

Fig. 8: MTF matrix table for one channel.

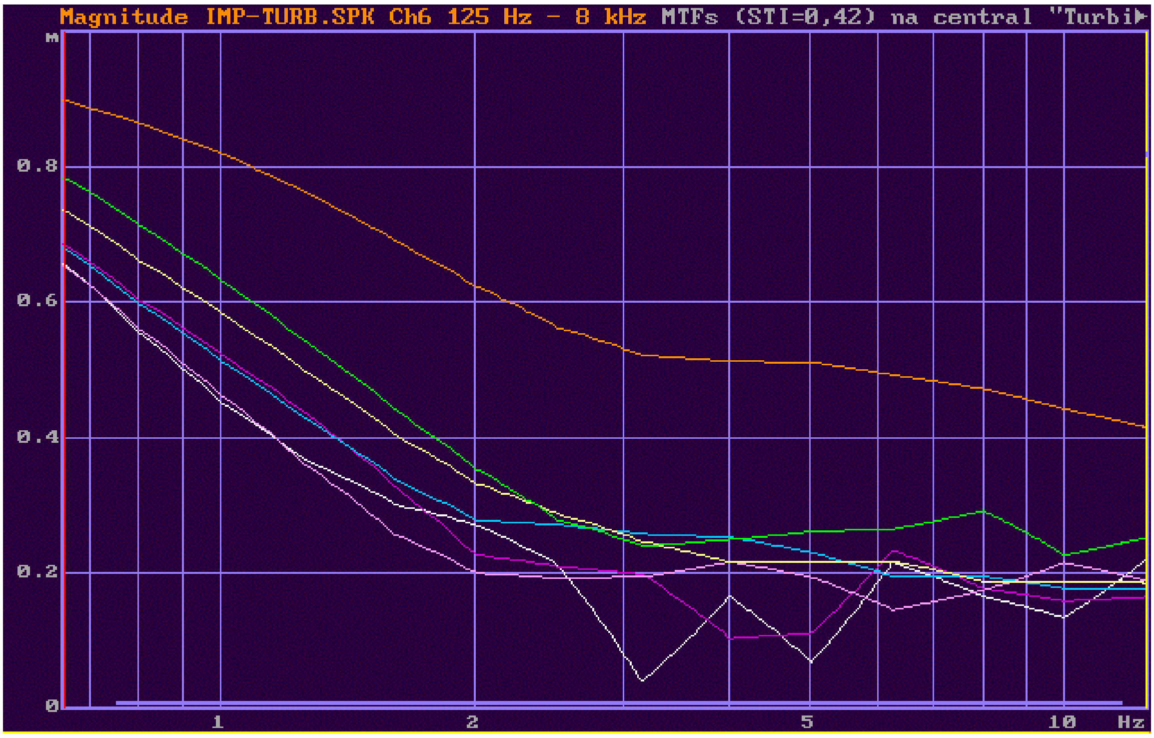


Fig. 9: MTFs available in spectral domain after pressing "Make SPK" in MTF matrix window.

1.1.7 Some basic mechanisms influencing the STI

In this chapter, the three basic types of disturbances that decrease the STI are presented as isolated theoretical examples: noise, reverberation, and echo. For these three examples, it is assumed that all other transmission properties are perfect. To simplify things, the corrections for masking and the absolute reception threshold are not taken into account. They would have very little influence (ca –0.02) anyway.

1.1.7.1 Noise

The influence of steady noise on the MTF is modulation-frequency independent. Together with the signal level, it can be calculated using the well-known formula

$$m_{SNR} = \frac{1}{1+10^{-\frac{S/R}{10}}}$$

An SNR of –15 dB or worse in all bands means that the final STI will be exactly 0. From –15 dB on, every increase by 3 dB of the SNR will raise the STI value by 0.1, until finally reaching 1 when the SNR rises to 15 dB or higher in all bands. The landmark 0 dB for the SNR will be translated into an STI of exactly 0.5.

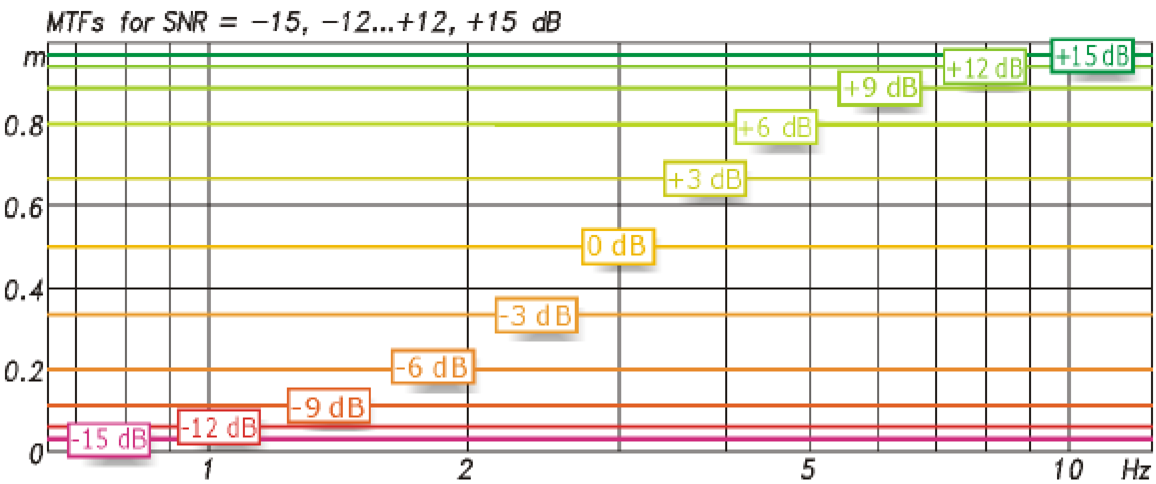


Fig. 10: Influence of SNR on the MTFs.

SNR [dB]	-15	-12	-9	-6	-3	0	+3	+6	+9	+12	+15
STI	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Tab 4: Relation between broadband SNR and STI.

1.1.7.2 Reverberation

Reverberation has a low-pass filter effect on the MTFs. With the assumption that the decay is strictly exponential, indicating perfect diffuse-field conditions, the influence of the reverberation time RT on the MTFs can be expressed as

$$m_{RT}(f) = \frac{1}{\sqrt{1+\left(\frac{2\pi \cdot f \cdot RT}{13.8}\right)^2}}$$

Fig. 11 shows the effect of various reverberation times RT on the MTF.

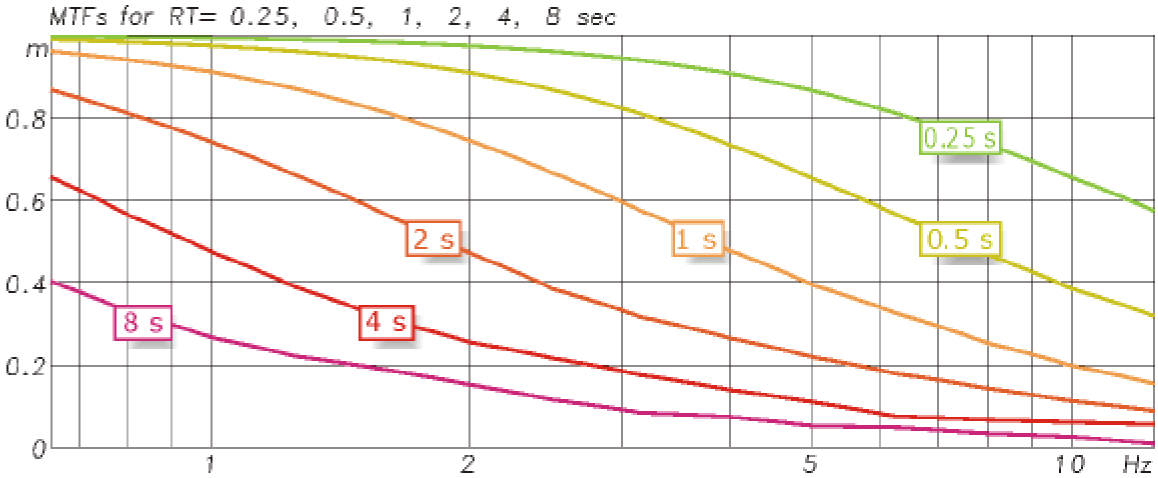


Fig. 11: Influence of different broadband RTs on the MTF.

RT [s]	8	4	2	1	0.5	0.25
STI	0.19	0.31	0.44	0.59	0.74	0.85

Tab 5: Relation between broadband RT and STI.

1.1.7.3 Echos

An echo has a comb-filter effect on the MTFs, causing beating through constructive and destructive interference depending on the modulation frequency. For a broadband echo, its level-dependent influence on the MTF can be calculated by

$$m_{ECHO}(f) = \frac{\sqrt{1+2\cdot I_E \cdot \cos(2\pi \cdot f \cdot T_E)+I_E^2}}{1+I_E} \quad \text{with} \quad I_E = 10^{\frac{L_E}{10}}$$

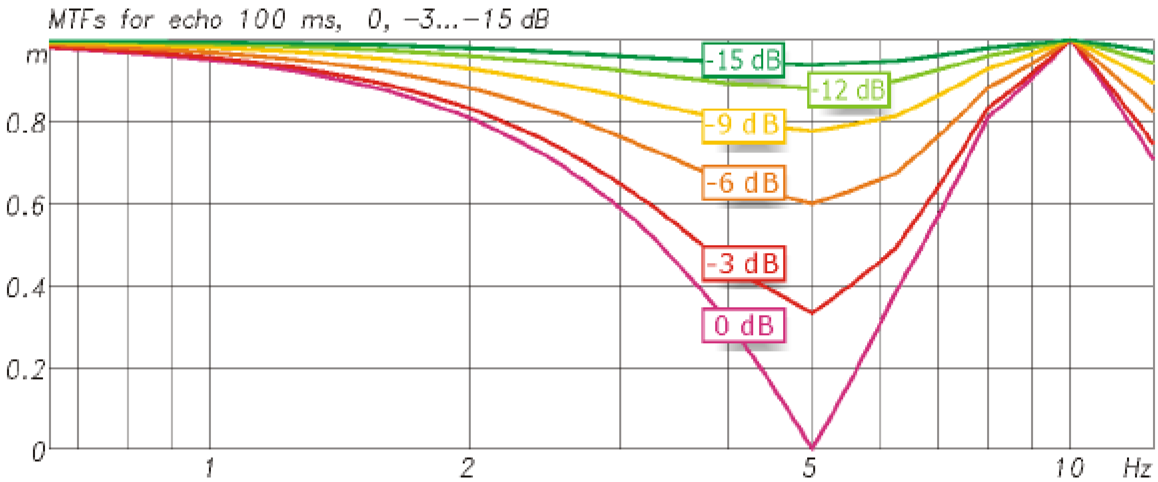


Fig. 12: Influence of the level of an echo with 100 ms delay on the MTFs.

L_E [dB]	0	-3	-6	-9	-12	-15
STI	0.68	0.74	0.80	0.87	0.93	0.97

Tab 6: Relation of 100 ms echo level and STI.

As can be seen, an echo of 100 ms arriving with the same level as the direct sound will cause the STI to only drop to 0.68, which would still be rated "good". However, the intelligibility will actually be poor in the presence of such a strong delayed component. This shows that the SNR is too insensitive to this kind of transmission disturbance. There are several other situations in which the STI fails to predict intelligibility, but they are out of the scope of this manual.