

Articulation- Prerequisite to Performance

by Arthur Noxon • Presented at the XXth AES Convention in NY, Nov 19XX

Summary

This paper explains the technical details behind what came to be known as the MATT Test. Acoustic Articulation goes up as the listener hears more signal and less noise. But how can we measure articulation in a real world audio situation and get an accurate result?

The Modulation Transfer Function is the established basis for testing the quality of speech intelligibility. This paper reviews the current of MTF test signals as the performance spec for hi-fi and pro sound playback rooms. Recordings of the test, made in a listening room under different conditions of acoustic treatment, will be played while hard copy is displayed.

Acoustic Articulation is the ability of an acoustic space to faithfully track signal level changes. That description alone is sufficient to warrant our attention to the subject. What would the world be like if we increased audio signal gain, but did not hear a corresponding sound level rise? What if we cut the signal power and did not experience a drop in sound level? Articulation is such a fundamental concept that it is easily taken for granted. It is the current best indicator for a communication channel and human perception. That is why we use articulation measurements as the baseline for evaluating sound systems.

Introduction

The search to define quality audio playback has for many years been keyed to electronic performance specifications. However, the final link in an audio chain is always the acoustic coupler, the interconnect between the speaker and the listener. The proverbial chain is still only as strong as its weakest link and with today's sophisticated electronics and transducers, the weakest link in the audio chain is undoubtedly the playback room. The question inevitably arises as to how to test the room as the final link in the audio chain and what should be the specification.

The long-standing test procedure for room acoustics is the RT-60 decay time measurement. In the last few years, a new acoustic test has been introduced into audio. It is the speech intelligibility test and it comes from the world of speech and communication. Intelligibility measurements combine the consequences of RT-60 with the room's background noise level to predict the integrity that remains of a modulated signal that has been transmitted across a room. This test is applied to the acoustic link of sound systems that are as huge as a dome stadium to as small as a telephone earpiece. Intelligibility testing is now beginning to impact pro sound and hi end audio, that is why it is the topic of this paper.

Over the last few years B & K (RASTI) and the Crown (Tecron) each have produced a procedure to measure speech intelligibility. Their data is converted into a single number, the STI (Speech Transmission Index). This test equipment only monitors the performance of an existing system and is not a piece of diagnostic equipment. The STI is a performance rating number, it does not help the engineer to know what to fix in order to get a better STI. The next generation of test equipment in this arena will naturally be of the diagnostic type.

The concern for intelligibility and how to measure it is not new. It dates back at least to early radio days with the problem of signal-to-noise ratio (SNR) that prevents messages from getting through. The development of the telegraph, telephone and radio, right on into today's deep space communications form a continuous chain of contributions to the advancement in the understanding of the perception of signals.

Speech Intelligibility

Within the last few years, Speech Intelligibility has surfaced as a performance requirement in sound systems. Engineers, designers, contractors and architects no longer only work towards smooth-sound level distributions and properly shaped octave band equalization (EQ) contours; now they are being required to meet Speech Transmission Index (STI) criteria. Speech intelligibility is a special application of the basic concept of articulation. It is a speech band limited and “weighted” version of articulation.

We encounter something similar when doing sound level measurements. The “A-Weighted” sound level frequency response curve is not a “flat” response curve, it has been modified to include the loss of efficiency of human perception in the lower and very high frequency range. It is the weighted response curve that is integrated over the audio range to achieve the total adjusted sound level in dB,A. This is directly analogous to the STI which is an integration of the articulation frequency response curve which has been weighted for the purpose of speech and communication.

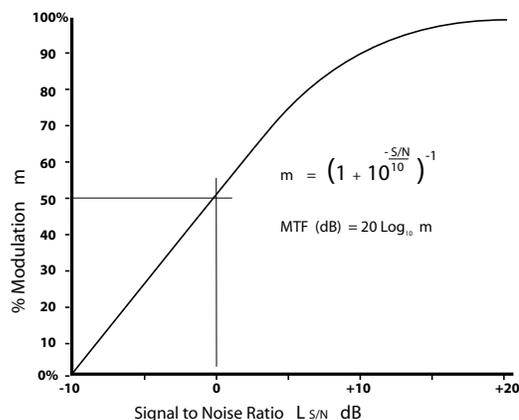
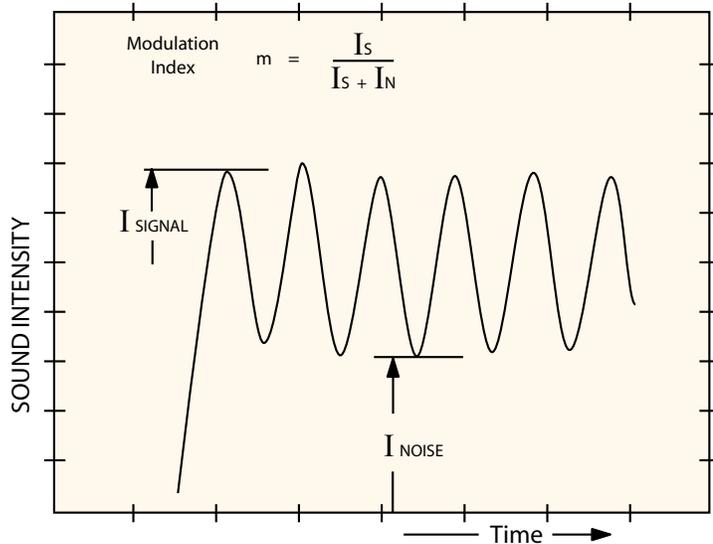
Modulation Transfer Function

The response curve that forms the basis of articulation measurements is called the MTF, or Modulation Transfer Function, ranges from zero to 100%. Zero percent MTF signifies that a modulated signal is undetectable by a person. Tone bursts, as in a Morse code transmission, would have absolutely no signal modulation at the receiving end. There are two ways this can happen.

To achieve zero signal modulation, the receiver could be a long way from the transmitter. It would receive nothing but background noise, “static” on the transmission channel. The tone sequence may well actually be received but it is not perceived by the listener if the signal is buried more than 10 dB below the background noise floor. The MTF is zero if the external noise is too loud compared to the modulated signal.

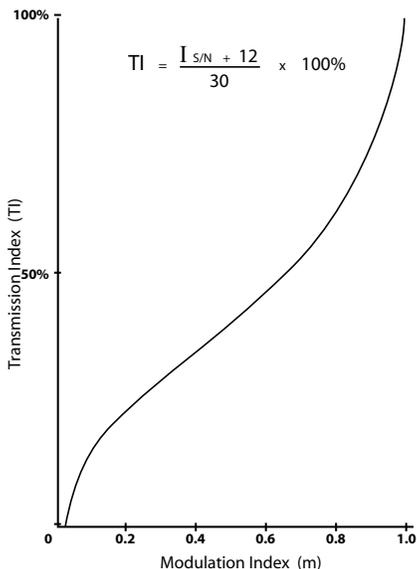
Another instance in which MTF drops to zero would occur when transmitting code across a reverb chamber. With a typical RT-60 of 10 seconds (sound level drops 60 dB in 10 seconds), the rapid staccato of a Morse code will be totally obscured by the room’s reverberant noise field. Because the tone of the reverberation sounds just like the signal, it masks the signal very easily. The reverberant field type of noise easily masks signal modulation that is 5 dB below the noise floor.

Fig. 1 Modulated Signal with Noise



The preferred signal perception is 100% MTF. Morse Code could easily have 40 dB of electronic signal modulation, the tone burst signal level relative to the circuit noise floor. People have limits to perceived modulation. Sound over 140 dB is painful and that under 10 dB is inaudible. Maximum perceptible modulation is 130 dB. That is why a 1000 dB signal-to-noise ratio is imperceptibly different from a 100dB SNR, assuming the signal strength for both signals was the same.

We might be able to tolerate 130 dB of signal level modulation but 20 dB has proven to be effectively full range. A 10 dB modulated SNR has proven clearly heard, this would occur if a 70 dB test



tone was placed in a 60 dB background noise level. The result of many studies in perception is that for effective communication, modulated 18 dB SNR is sufficient to be called 100% modulation. At the other end is ½ dB modulation which is essentially imperceptible. The dynamic range for modulated signals that is significant to human perception is about 18 dB. With these two end points defined, all that remains is to fill in between the intervening points. Much research into human perception has been spent in developing this relationship shown in Figure 1.

Signal to Noise Ratio

By now it should be clear that an articulation test measures both the dynamic and static behavior of sound levels. A third-octave or other RTA device measures static sound level conditions. The sound levels of a facility can be measured first without and later with a signal applied and the MTF can be evaluated with respect to background noise.

The background noise spectrum can be loaded into “Memory A” of an RTA. Then power up the sound system and measure pink noise levels at the listening position. Load them into “Memory B.” The difference between these two curves is the SNR vs. frequency curve. An example of this is shown in Figure 2.

The SNR can be converted to MTF by using Figure 1. The resulting TI (Transmission Index) vs. frequency curve of Figure 4 is a linear, unweighted response curve. For speech intelligibility the TI is multiplied by the weighting curve for (Figure 1) speech. The result shown in Figure 5 is the band-limited STF (Speech Transfer Function) curve. The percent of the area coverage under the STF equals the STI, Speech Transmission Index.

Fig. 2 PA & Crowd Noise Levels

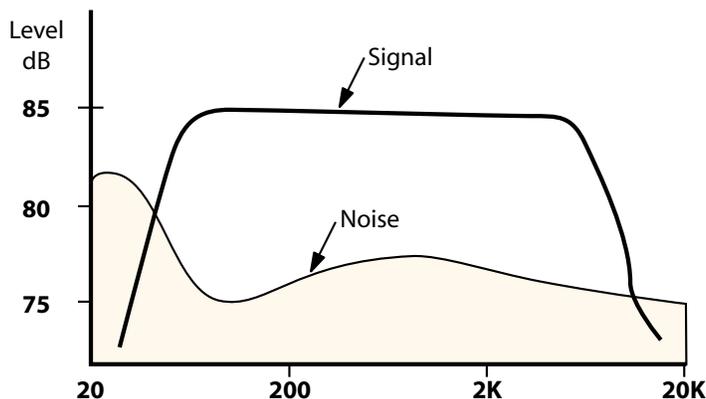
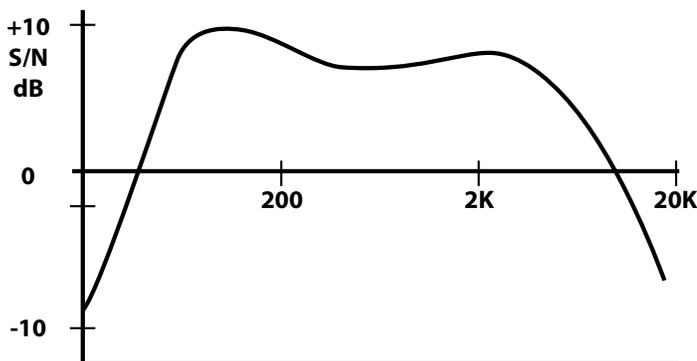


Fig. 3 Signal to Noise Ratio

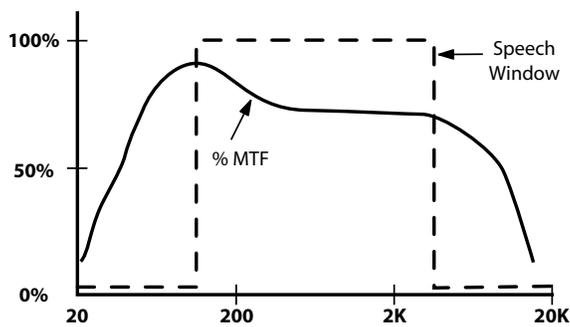


This signal to background noise version of MTF analysis is fairly straight forward. Most of us in audio could produce today the STI by using an RTA, the MTF-S/N chart, the STF weighting curve and a lot of data plotting. This version of MTF has limited application. Conceptually, it measures the quality of communication for an anechoic chamber filled with background noise the announce system in a noisy, large factory or the PA for a huge, noise crowd of people might be a reasonable application.

Signal to RT-60 Ratio

The other aspect of MTF includes reverberation, the more common problem in audio playback. Reverberation is the energy that lingers after a signal has been transmitted. No matter how reverberant a space may be, the residual energy will eventually die away leaving the ambient background noise as the sound in the room. If an alarm went off every hour in a reverb chamber a valid signal would be received because the time between signals far exceeds the decay time of the reverb chamber. Conversely, a high-speed Morse code transmitting four bursts per second would be converted to a total blur of noise, completely inaudible signal modulation.

Fig. 4 TI & Speech Window



As a consequence of reverberation, the signal modulation rate or bursts per second is related to the MTF. Slow burst rates naturally have good MTF and fast burst rates often have poor MTF. The range of burst rates that matter to people and

communication is the range from 2 Hz to 20 Hz and the MTF vs. Reverberation, shown in Figure 6. Burst rates above 20 Hz sound like a low frequency note and therefore are not capable of being a modulate signal.

Real World MTF

The two basic versions of signal-to-noise have been presented. Background noise and reverberation are combined in most real-life situations. If the MTF for these two independent processes can be determined and the combined effect is desired, then we multiply the background noise MTF by the RT-60 MTF. The result gives the combined effect of substantial background noise in a reverberant space.

For example, consider a noise basketball game in a gymnasium. The crowd noise level could be 85 dB,A. The PA might be set at 90 dB. The RT-60 of the occupied gym might be 2.5 seconds. Shown in Figure 7 the SNR of 5 dB gives 75% partial MTF due to the PA level and crowd noise. The MTF/RT-60 curve gives a partial MTF of about 50% due to the gym reverberance at 2 bursts/sec. The combined effect is a MTF of about 35%, pretty bad. Successful announcers instinctively understand this and enunciate slowly to utilize the intelligibility benefits that go with slow modulation rates.

3-Dimensional MTF Displays

With MTF, the signal modulation rate is not impacted by the background noise levels but it is strongly effected by the RT-60. Low modulation rates are more audible than fast modulations in a reverberant space. At the lowest modulation rate, the MTF is usually controlled by the background or external noise. MTF for the higher burst rates are controlled by the reverberation of the room.

The full audio frequency ranges from 20 Hz to 20 KHz. Not only does the background noise spectrum vary with frequency, the RT-60 will also vary with frequency. The next step then is to perform the MTF analysis throughout the full frequency range. The MTF frequency response curve is absolutely essential for a detailed analysis or diagnostics of the communication channel.

If both the modulation and tonal frequency aspects of MTF are combined, the result appears as a 3-dimensional print-out, or

Fig. 5 Speech Transfer Function (STF)

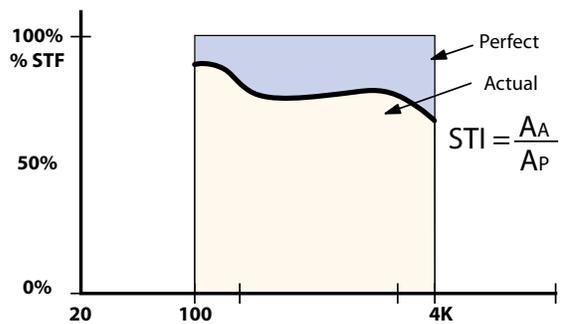


Fig. 7 MTF: RT-60 with Noise, Example

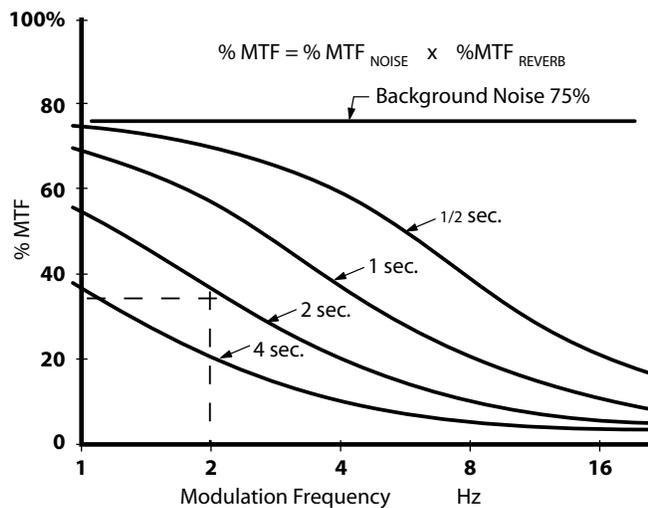
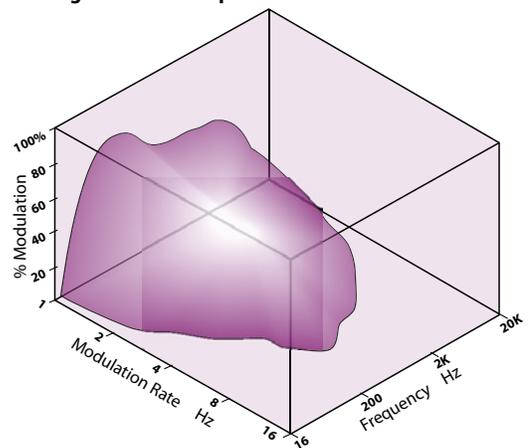


Fig. 8 MTF: 3-D Space



the MTF waterfall. Figure 8 illustrates this display. The present day's use of MTF analysis is dedicated to speech intelligibility. It is limited (Figure 9) to modulation rates between 2 and 8 Hz, and a frequency range between 100 Hz and 4 KHz. This is 1/6 of the total 3-dimensional MTF volume available to human perception. Depending on the application, different sections of the MTF volume will be used. For example, as shown in Figure 10, a Morse Code transmission would need a narrow range, about 1/30 of the total MTF space.

A typical recording studio control room and quality hi-fi listening room are required to handle a wide frequency range and be capable of fast modulation rates. Figure 10 also shows how a precision playback room might occupy 50% of the full MTF space. Dynamic stability might be required up to a 12 Hz modulation rate for any frequency ranging between 40 Hz and 16 KHz.

A digital sampling studio could have even higher expectations and be required to track well into the first 70% of the MTF space. It might have the full frequency bandwidth of 20 Hz to 20 KHz and handle up to a 15 Hz modulation rate. The MTF volume for various categories of performance can only be estimated at this time as they have yet to be properly defined.

Conclusion

The role of MTF analysis in audio is just beginning to make its presence felt. For the last two years it has been making its way into audio by the way of commercial sound systems. An advancement into one specialty area of audio eventually makes its presence felt in all areas of audio. It is safe to expect that in the next decade we will be using another rackmount, the MTF will probably be located just above the RTA and EQ. There can be no doubt that by including human perception of signals as an audio performance indicator we will produce even better, more accurate and most importantly, more relevant audio playback systems.

Fig. 9 Speech MTF Volume

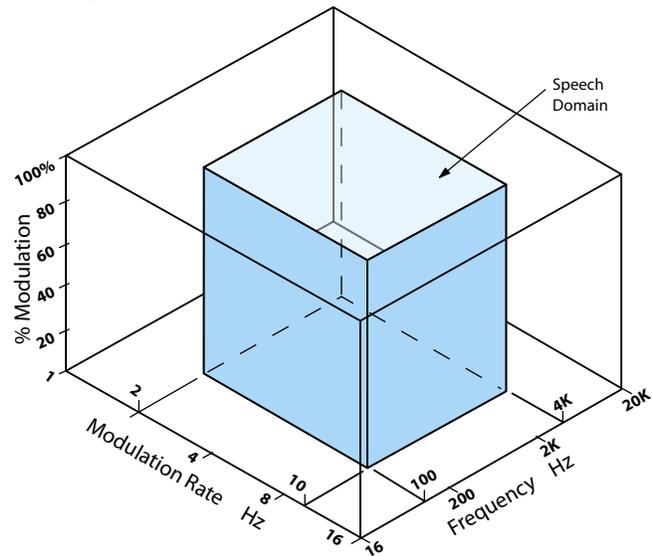


Fig. 10 Specialty MTF Volumes

