Measuring Speech Intelligibility with the SENCORE SoundPro Audio Integrator

What did he say? I couldn’t understand him! We have all experienced difficulty understanding a speaker or message over a P.A. (public address) system. Why? Why is a speaker or P.A. message so clear sometimes and so difficult or impossible to understand in others?

There is a wide range of issues involved with the science of speech intelligibility or the ability of a listener to understand a spoken word or phrase. While it is a complex issue, this article attempts to summarize the subject and offer technical solutions to quantifying speech intelligibility with electronic measurement.

Understanding Human Speech

A normal speech level is between 55 and 65 dB at approximately 3 ft. from the speaker. Speech levels may dip to 30 dB when whispering or peak over 90dB when shouting. It should be noted that above 75 dB, human speech becomes more difficult to understand due to voice phonetic changes and overloading of the ear.

The fundamental sound waveform of human speech ranges from 100-400 Hz. Men average around 100Hz while women average near 200 Hz. The resonant characteristics of our vocal tract result in various level harmonic content of the fundamental waveform. Words or phrases are made-up of spoken vowel and consonant sounds. The resonances as we transition and say various vowel sounds give us a distinct and recognizable voice.

Sound energy or level is mostly in the longer duration vowel sounds lasing 30-300 mS. Word recognition, intelligibility, is mostly determined by the shorter duration consonant sounds lasting
only 10-100 mS. Consonant sounds are fast duration sounds, somewhat like impulses or noise, producing higher sound frequencies in the range of 2 kHz to 9 kHz. Since consonant sounds are faster, much lower in sound level (10-30 dB), and occupy higher frequency bands, they are also the most difficult to hear and discern. Thus, the CHARLIE BROWN in the classroom effect!

Complicating matters is the fact that spoken sounds are increases and decreases in frequency content at various rates. The sounds merge and overlap in the time domain. By analyzing the spectrum content of a speech sequence, it can be shown that the individual frequency or audio octave bands vary content at rates ranging from 0.2 to 12 times a second. This is often referred to as modulation. Speech Intelligibility depends on preserving the integrity of these changes or modulations of the spoken voice as discerned by the listener.

![Speech Audio Octave Bands](image)

**Figure 2. Speech Audio Octave Bands. Frequency content varies (modulates) in each octave band at rates ranging from 0.2 to 12 times a second.**

**Speech Comprehension – Preserving Sound**

While the precise ear-brain mechanism of decoding speech may never be fully understood, many factors are known to influence it. Perhaps speech comprehension is best understood by understanding how to preserve it. First, preservation of the speech spectral energy, both in level and frequency content, is crucial. Second, the sound must be delivered to a listener without competition of interfering sounds or noise arriving at the same time, which masks comprehension. Finally, room reverberations cause sounds to arrive at the listener at delayed times, arriving with the direct sound energy causing difficulty in discerning speech sounds.

Preserving the frequency or spectral content of speech is important when sound systems are being used. The frequency response of microphones, amplifiers and speakers must have a wide response. Good systems cover a frequency range of 80 Hz to near 10 kHz for good high frequency consonant reproduction. A poor response with a high frequency roll-off at 2 kHz would make it difficult to discern between “f” and “s” or “d” and “t” sounds. A relatively flat response is desirable because a peak or dip in frequency response can degrade the ability to discern sounds. The frequency response must be consistent throughout the listening area, demanding good speaker coverage with minimal high frequency roll-off.
Preserving the frequency or spectral content requires that a minimal amount of distortions exist in the sound system. Any clipping or inter-modulation distortion (IMD) products result in harmonic signal content (clipping) and mixing products (IMD sum & difference signals) not contained in the original sound. As severity increases, the spectral content interferes with sound comprehension. Microphones and amplifiers are prone to clipping, while some loudspeakers are prone to IMD at higher volume levels.

Probably the biggest threat to speech comprehension is competing noise, voices or other sounds reaching the listener. Sounds can be room noises, crowd noises, electronic system noises, etc. The added sounds arriving to the listener position compete with speech sounds, making ear/brain interpretation difficult. The amount, noise level, and frequency content of the noise, determines the severity which is expressed as a signal-to-noise dB ratio. Typically, the speech sound must be 12 dB louder than the noise for good word recognition. Studies have shown that 2-3 competing voices offer just as much interference as broadband noise.
Similarly threatening to speech comprehension is the reverberation effects of the room or space. Reverberation causes the same sound energy spectrum to reach the listener multiple times through various delays. Short delays, less than 50 mS, actually enhance speech perception. Longer delays essentially cause competitive noise to the direct sound. In large cathedrals or sports arenas the reverberant sound energy arrives from all angles multiple times which is highly destructive to speech comprehension.

**Measuring Speech Intelligibility**

Although hearing is a complex matter, the ability to hear and understand speech produced by a speaker or sound system in a particular setting can be statistically graded or valued. In general, there are two forms of speech intelligibility testing in which to derive a value that relates to speech intelligibility.

1. Human Based Testing
2. Electronic Testing Based

With human based testing listeners monitor specially constructed speech samples broadcast over the sound system and report the words and sentences they hear. Statistical results then are compiled to determine how much of what was said was heard by the listeners. While considered the most accurate, human based testing is not practical for most applications as it is expensive and time consuming.

Electronic based testing offers a more feasible alternative to human based speech intelligibility testing. Electronic testing involves introducing sound energy into the room and analyzing the room’s noise and/or reverberation effects on the introduced sound. Because the effects of noise and reverberation on speech intelligibility are largely understood, electronic analysis can produce ratings or values that closely agree with human based speech intelligibility testing.

Electronic measurement of speech intelligibility has evolved in understanding and testing sophistication. While many test methods have been used in the past, there are currently 3 common testing methods used to gauge speech intelligibility using electronic analysis. They include:

1. % ALCONS (Percentage Articulation Loss of Consonants)
2. RASTI (Rapid Speech Transmission Index)
3. STI-PA (Speech Transmission Index for Public Address Systems)

Electronic testing involves introducing a test signal into the room and analyzing the signal at the listening position with an electronic measurement instrument such as the Sencore “SoundPro.” The Sencore SP395 SoundPro performs all three of the speech intelligibility tests. The more popular RASTI and STI-PA tests are explained in the following sections.
Speech Intelligibility – RaSTI

Rapid Speech Transmission Index (RaSTI) offers a simple, less time-consuming approach to speech intelligibility testing. RaSTI calculates an index value (0.00 to 1.00, with 1.00 being perfect intelligibility) by evaluating 9 modulation waveforms, four in the 500 Hz and five in the 2 kHz octave bands.

RaSTI measurements use a transmitter to output a composite signal of all 9 modulation waveforms, and a separate receiver which analyzes the composite waveform. The SoundPro analyzes the results of a standard RaSTI test signal, which is supplied on a CD. By using a separate test signal on a CD, the test signal can be routed to a system that has a number of speakers such as an emergency evacuation system, and the SoundPro can be carried around to measure each speaker individually.

In operation, an amplifier and speaker are set up at the “speaker” position, and the SoundPro microphone is set up at the “listener” position. The RaSTI test waveform is played from a CD or other source, and the test is performed by the SoundPro RASTI analyzing function. After 10 seconds of analysis, the results are shown on the screen, and the RaSTI value is read directly. See table 1 showing the RaSTI measured values and a summary of the results of these tests.
RaSTI Function Operation

1. **Apply a test signal to the system.** Connect an amplified speaker to a CD player and insert the RASTI test signal CD. The location of the speaker is the “speaker” location.
2. **Microphone** - Set up the SoundPro microphone at the “listener” position. Use a stand, if possible.
3. **Level** - Using the SPL meter function, set the level of the test signal as measured by the SoundPro to the “nominal” listening level. In most cases this will be 75-85dB SPL.
4. **RASTI** - Select the RASTI function on the SoundPro.
5. **Begin Test** Click on the ON/OFF field changing the “Off” to “Run”. This begins the test.
6. **Wait For test to complete** - Allow at least 12 seconds for the analysis to be completed.
7. **RaSTI Result** - Read the RaSTI value on the bottom of the screen.

Speech Intelligibility Results

Use this table as a guideline for evaluating the results of the RaSTI speech intelligibility tests. A value of 0.2 to 0.4 is poor, 0.4 to 0.6 is fair, and above 0.65 is good.

<table>
<thead>
<tr>
<th>% Alcons</th>
<th>RaSTI</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.70</td>
<td>0.20</td>
<td>Bad</td>
</tr>
<tr>
<td>44.00</td>
<td>0.25</td>
<td>Bad</td>
</tr>
<tr>
<td>33.56</td>
<td>0.30</td>
<td>Poor</td>
</tr>
<tr>
<td>25.59</td>
<td>0.35</td>
<td>Poor</td>
</tr>
<tr>
<td>19.52</td>
<td>0.40</td>
<td>Poor</td>
</tr>
<tr>
<td>14.89</td>
<td>0.45</td>
<td>Fair</td>
</tr>
<tr>
<td>11.35</td>
<td>0.50</td>
<td>Fair</td>
</tr>
<tr>
<td>6.60</td>
<td>0.60</td>
<td>Fair</td>
</tr>
<tr>
<td>5.40</td>
<td>0.65</td>
<td>Good</td>
</tr>
<tr>
<td>5.84</td>
<td>0.70</td>
<td>Good</td>
</tr>
<tr>
<td>2.93</td>
<td>0.75</td>
<td>Good</td>
</tr>
<tr>
<td>2.23</td>
<td>0.80</td>
<td>Excellent</td>
</tr>
<tr>
<td>1.70</td>
<td>0.85</td>
<td>Excellent</td>
</tr>
<tr>
<td>1.30</td>
<td>0.90</td>
<td>Excellent</td>
</tr>
<tr>
<td>0.99</td>
<td>0.95</td>
<td>Excellent</td>
</tr>
<tr>
<td>0.76</td>
<td>1.00</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

Table 1. % Alcons and RaSTI values and related test ratings.
Speech Intelligibility – STI-PA

STI-PA spans the 125 Hz to 8000 Hz octave bands using one modulation speed for each data point. The SoundPro STI-PA implementation is unique, using an STI-PA excitation signal and proprietary DSP algorithms for determining loss of modulation. The STI-PA test signal, which sounds like a steam engine train, is played from the supplied CD into the microphone, and the signal is sampled at the listening position for 15 seconds. The STI-PA value is calculated and shown on the SoundPro test screen.

Figure 7. The SoundPro STI-PA measurement screen and control field descriptions.

STI-PA Function Operation

To use STI-PA, you need to play the STI-PA test signal through a speaker, and then use the SoundPro microphone to measure the STI-PA signal in the room. If you are testing a specific installed system, such as a fire evacuation sound system, send the signal through the system amplifier and speaker.

1. **Set up the connections.** Connect the STI-PA test signal to the speaker system. Play the test signal CD and adjust for appropriate level. In some cases, a test level SPL will be specified. Otherwise, typically a nominal level such as 70-80 dB SPL is used.

2. **Start the test.** Click on the Run field to begin the test. Wait for 15 seconds for the test to complete. While the test is running, the test microphone should remain in the same position, and the noise level in the room should remain stable.

3. **Read the results.** After the test is complete, you will see a set of signal to noise (s/n) and m values. Also, the final STI value is shown. If a condition was detected that prevents getting an accurate value, you may see the message “low input” or “overload” on the screen, and the data values are not shown.

Summary

Electronic speech intelligibility tests determine if rooms and public address systems can adequately transmit speech that can be understood. Tests accurately quantify the performance of the sound system and environment in which they operate. Results clearly identify if improvement is needed or if performance criteria is met.
Speech Intelligibility tests are available as an option to the Sencore SoundPro Audio Integrator, model SP395. For more information on speech intelligibility testing or the SP395 SoundPro call 1-800-SENCORE (736-2673) or visit http://www.sencore.com.


http://www.sencore.com  mailto:sales@sencore.com  1.800.736.2673 or 1.605.339.0100