



Effects of Physical Environment on Speech Intelligibility in Teleconferencing

(This article was published at Sound and Video Contractors website www.svconline.com in June 2005)

By Mei Wu and James Black

We probably all have experienced poor speech intelligibility during teleconferencing. It happens even when high quality sound systems are used at both talker and listener's rooms and even when both rooms have good speech intelligibility during local conferences. These observations reveal two facts (1) physical environment in conference rooms is as important as a sound system and (2) physical environment requirements for teleconferencing are more stringent than for local conferencing. This article shows how physical environment affects speech intelligibility in teleconferencing. A bad environment can spoil speech intelligibility despite the merit of a sound system, and a good environment can enable a sound system to achieve its design potential.

Assessing Speech Intelligibility with STI

Speech intelligibility is determined by measuring the proportion of test items, such as words or syllables, that are heard correctly. In a typical speech intelligibility test, a specified set of syllables, words, phrases or sentences is presented to a listener. The listener responds by writing down what was heard. Test results collected over the years show that speech intelligibility is reduced by the increase of background noise (or decrease of signal-to-noise ratio) and by increase of reverberation time. Typical curves showing the relationship between speech intelligibility and signal-to-noise ratio or reverberation time are shown in Figures 1 and 2.

Instead of going through speech intelligibility tests, indices have been developed to assess speech intelligibility based on measurements of physical environment, i.e. background noise and reverberation time. The commonly used indices include Speech Interference Level (SIL), Articulation Index (AI), Speech Intelligibility Index (SII) and Speech Transmission Index (STI). The authors prefer to use STI because the first three indices calculate speech intelligibility only from signal-to-noise ratio (with modifications for reverberation time and other factors). STI is the only method which takes into account both signal-to-noise ratio and reverberation time.

Speech Transmission Index (STI) was developed in the Netherlands by Tammo Houtgast and Herman Steeneken¹. It determines speech intelligibility based on the modulation depth of a

¹ Houtgast, T. and Steeneken, H. J. M. "Evaluation of Speech Transmission Channels by Using Artificial Signals." *Acoustica*, vol. 25, pp 355-367, 1971 and "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics," *Acoustica*, vol. 46, pp 60-72, 1980.

speech waveform. Figure 3 is a typical speech waveform showing instantaneous sound pressure as a function of time. It has a high frequency carrying wave with its amplitude modulated by a low frequency modulation wave. The resolution of the time scale in Figure 3 is such that the carrying wave itself cannot be seen in any detail, but the amplitude modulation is clearly evident. The amplitude-modulated speech waveform in Figure 3 shows major peaks at roughly 100, 300, 600, 850 and 1100 milliseconds. The difference in level between a peak and an adjoining valley is referred to as the depth of modulation. If no noise or reverberation alters speech, there is very little energy in the valleys between peaks, modulation depth is 100%, STI value is 1 and speech intelligibility is excellent. Background noise and/or reverberation add energy in the valleys, reduce the depth of modulation, and reduce STI value and speech intelligibility. When STI value goes down to 0, speech is totally unintelligible. The table below shows STI values and corresponding speech intelligibility assessment.

STI and Speech Intelligibility

STI Value	Speech Intelligibility Assessment
0.75 <= STI	Excellent
0.60 <= STI < 0.75	Good
0.45 <= STI < 0.60	Fair
0.30 <= STI < 0.45	Poor
STI <= 0.30	Bad

In the complete version of STI testing, the modulation depth is measured in 98 tests over 14 modulation frequencies (0.63 Hz to 12.5 Hz in 1/3 octave bands) for 7 octave bands (from 125 Hz to 8,000 Hz) of carrying wave. Simplified Speech Transmission Index measurement methods, such as RASTI (Rapid Analysis – or Room Acoustics as known by some – Speech Transmission Index) and STIPA (Speech Transmission Index – Public Address), are defined in Standard IEC 60268-16.

STI can also be determined from calculated modulation transfer function, when the impulse response of a room can be regarded as a well-behaved room-response with an exponential decaying envelope. A simplified formula for modulation transfer function at frequency F can be expressed as a function of the reverberation time T and effective signal-to-noise ratio S/N:

$$m(F) = \frac{1}{\sqrt{1 + \left(2\pi F \frac{T}{13.8}\right)^2}} \cdot \frac{1}{1 + 10^{(-S/N)/10}}$$

Case Studies

During teleconferencing, when speech in a talker’s room is transmitted to a listener’s room, background noise in the talker’s room is also transmitted and amplified in the listener’s room. While direct sound is transmitted to the listener’s room, reflected sound is also transmitted to the listener’s room. These reduce the signal-to-noise ratio and extend the reverberation time, and consequently reduce speech intelligibility. As a result, even if the sound system is perfect,

speech intelligibility during teleconferencing is lower than during local conferencing. The authors' experience is that conference rooms with good to fair speech intelligibility for local conferencing may have fair to poor speech intelligibility in teleconferencing. Therefore, before installing a teleconferencing system, it is sensible to consult a professional to ensure that the physical environment in a conference room will sustain a good to fair speech intelligibility.

The following case studies demonstrate how physical environment, such as background noise, reverberation time, microphone distance and orientation, affects speech intelligibility. To concentrate on the effects of physical environment, we assume the sound systems are perfect. The physical conditions of each case are listed in tables. Column 1 lists the case numbers. Column 2 is the tested or predicted STI values. Column 3 gives a description of speech intelligibility. Column 4 shows either the STI is in the talker's room or the listener's room. Column 5 shows the distance from the talker to the receiver (a microphone or listener in the talker's room). Column 9 gives a brief description of the rooms. The last column is the distance between a loudspeaker and listeners in the listener's room.

Cases 1 and 2 are STI values measured in a 12x24 square feet conference room with acoustical ceiling, gypsum board walls and carpeted floor. We can see that for local conferencing, speech intelligibility is good for listeners at 45 degrees 3 feet from a talker with a normal voice (Case 1) and at 60 degrees 6 feet from a talker with a normal voice (Case 2). When the speech is transmitted through a perfect sound system to another room, with the talker using a raised voice, speech intelligibility reduces to fair and poor² (Cases 3 and 4).

Case Number	STI	Description	STI Position (Local = 0, Tele. = 1)	Talker Distance to Receiver (Mic) [ft]	Vocal Effort (Normal = 0, Raised = 1)	Directivity (deg from axis in Hor. Plane)	Directivity (deg from axis in Ver. Plane)	Room Description
1	0.69	Good	0	3	0	45	0	conf. rm.
2	0.66	Good	0	6	0	60	0	conf. rm.
3	0.55	Fair	1	3	1	0	0	conf. rm.
4	0.44	Poor	1	6	1	0	0	conf. rm.

Comparing Case 4 with Case 3 we can see that when the microphone is moved closer to the talker (from 6 feet to 3 feet), speech intelligibility improves from poor to fair.

Cases 5 through 9 are telemedicine cases. The talker's room is a 20x30x10(h) cubic feet operating room with gypsum board ceiling and walls, tile floor and 8 people. The microphone is located 3 feet above the surgeon's head at about 60 degrees. The background noise level in the operating room is about NC-45³. The noise is mostly generated by the air handling system and

² To simplify the analysis, we assumed the listener's room was acoustically identical to the talker's room and simulated the effects of background noise and reverberation time in the listener's room.

³ NC stands for Noise Criterion, which is a single value index commonly used to quantify the background noise level generated by air handling systems. The NC rating at a location is determined by comparing the octave-band sound pressure level spectrum measured at the location with the standard Noise Criterion (NC) curves. The lowest NC curve above the measured sound pressure level spectrum sets the NC level at the location. Noisy spaces have high NC values. The usually recommended NC rating for an open plan office is NC-40.

surgery equipment. A group of 10 people are observing the surgery in an observation room (listener's room). The room is 20x15x10(h) cubic feet with gypsum and glass walls with acoustical ceiling and carpeted floor. The voices are transmitted from the operating room to the observation room through a perfect sound system. The loudspeaker in the observation room is located in the front of the observers, facing them. The background noise in the observation room is about NC-40.

Case Number	STI	Description	Calculate STI Position (Local = 0, Tele. = 1)	Talker Distance to Receiver (Mic) [ft]	Vocal Effort (Normal = 0, Raised = 1)	Directivity (deg from axis in Hor. Plane)	Directivity (deg from axis in Ver. Plane)	Room Description	Speaker Distance to Receiver [ft]
5	0.40	Poor	0	2	0	0	0	operating	
6	0.24	Bad	0	4	0	90	0	operating	
7	0.31	Poor	1	3	0	0	60	op / obs	10
8	0.36	Poor	1	3	1	0	60	op / obs	10
9	0.37	Poor	1	3	1	0	60	op / obs	5

Case 5 shows that in the operating room, when the surgeon speaks, the speech intelligibility is poor (STI 0.40) to a person at 2 feet in front of him/her. Case 6 shows that speech intelligibility is bad (STI 0.24) to the person standing at 90 degrees from the direction of surgeon's speech at the end of the operating table. Case 7 shows that speech intelligibility to the people in the observation room is also poor (STI 0.31). Case 8 shows that if the surgeon raises his/her voice, speech intelligibility improves some but is still poor (STI 0.36). Case 9 shows that moving the loudspeaker closer to the observers does not help much (STI 0.37).

These STI values are expected considering the fact that the operating room has no sound absorptive material and has a reverberation time of 2.4 seconds. In a room with such a long reverberation time, people may be able to understand simple orders, but cannot carry a conversation. Traditional speech intelligibility vs. reverberation time curves shown in Figure 2 also indicate a poor speech intelligibility (Percent Intelligibility below 35%) in a room with 2.4 second reverberation time.

Case 10 shows how adding sound absorptive materials on the ceiling of the operating room improves speech intelligibility from poor (STI 0.36 in Case 8) to fair (STI 0.52). Here we assume the sound absorbing materials meets the cleanability, bacteria resistance and low particle shedding requirements for operating rooms.

Case Number	STI	Description	Calculate STI Position (Local = 0, Tele. = 1)	Talker Distance to Receiver (Mic) [ft]	Vocal Effort (Normal = 0, Raised = 1)	Directivity (deg from axis in Hor. Plane)	Directivity (deg from axis in Ver. Plane)	Room Description	Speaker Distance to Receiver [ft]
8	0.36	Poor	1	3	1	0	60	Operating	10
10	0.52	Fair	1	3	1	0	60	A. Ceil	10

Case 11 shows how adding 70 square feet of sound absorptive materials⁴ to the walls of the conference rooms in Case 3 improves speech intelligibility from fair (STI 0.55) to good (STI 0.62).

⁴ We assume that the sound absorptive materials are used properly to avoid flutter echo, enhance early reflect, etc.

Case Number	STI	Description	Calculate STI Position (Local = 0, Tele. = 1)	Talker Distance to Receiver (Mic) [ft]	Vocal Effort (Normal = 0, Raised = 1)	Directivity (deg from axis in Hor. Plane)	Directivity (deg from axis in Ver. Plane)	Room Description	Speaker Distance to Receiver [ft]
3	0.55	Fair	1	3	1	0	0	conf. rm.	3
11	0.62	Good	1	3	1	0	0	70 s.ft	3

Cases 12 and 13 show test results of how reducing the background noise from 45 dBA to 35 dBA in a 25x25 sq. ft. conference room improves speech intelligibility from fair to good, although we realize that in some cases reducing background noise may not always be a valid option, if the noise is mostly audience noise.

Case Number	STI	Description	STI Position (Local = 0, Tele. = 1)	Talker Distance to Receiver (Mic) [ft]	Vocal Effort (Normal = 0, Raised = 1)	Directivity (deg from axis in Hor. Plane)	Directivity (deg from axis in Ver. Plane)	Ambient Noise	Speaker Distance to Receiver [ft]
12	0.54	Fair	0	6	0	90	0	45 dBA	
13	0.65	Good	0	6	0	90	0	35 dBA	

Final Notes

The case studies presented in the article demonstrate how physical environment, such as acoustical treatment, background noise and microphone location and orientation, affects speech intelligibility during teleconferencing. By improving the physical environment, without changing the sound system, speech intelligibility can be improved from poor to good. It should be cautioned, however, that the calculations listed in the tables should not be used as general rules to predict the performance of any conference rooms because STI values vary with acoustical conditions. It is recommended that professionals be consulted and STI calculations be conducted for important teleconferencing rooms to ensure that the physical environment will sustain the performance of the sound system.

* * *

Mei Wu and James Black are acoustical consultants at Mei Wu Acoustics. Their resumes can be found at www.mei-wu.com. For questions or further information on speech intelligibility please contact Mei Wu at meiwu@mei-wu.com.

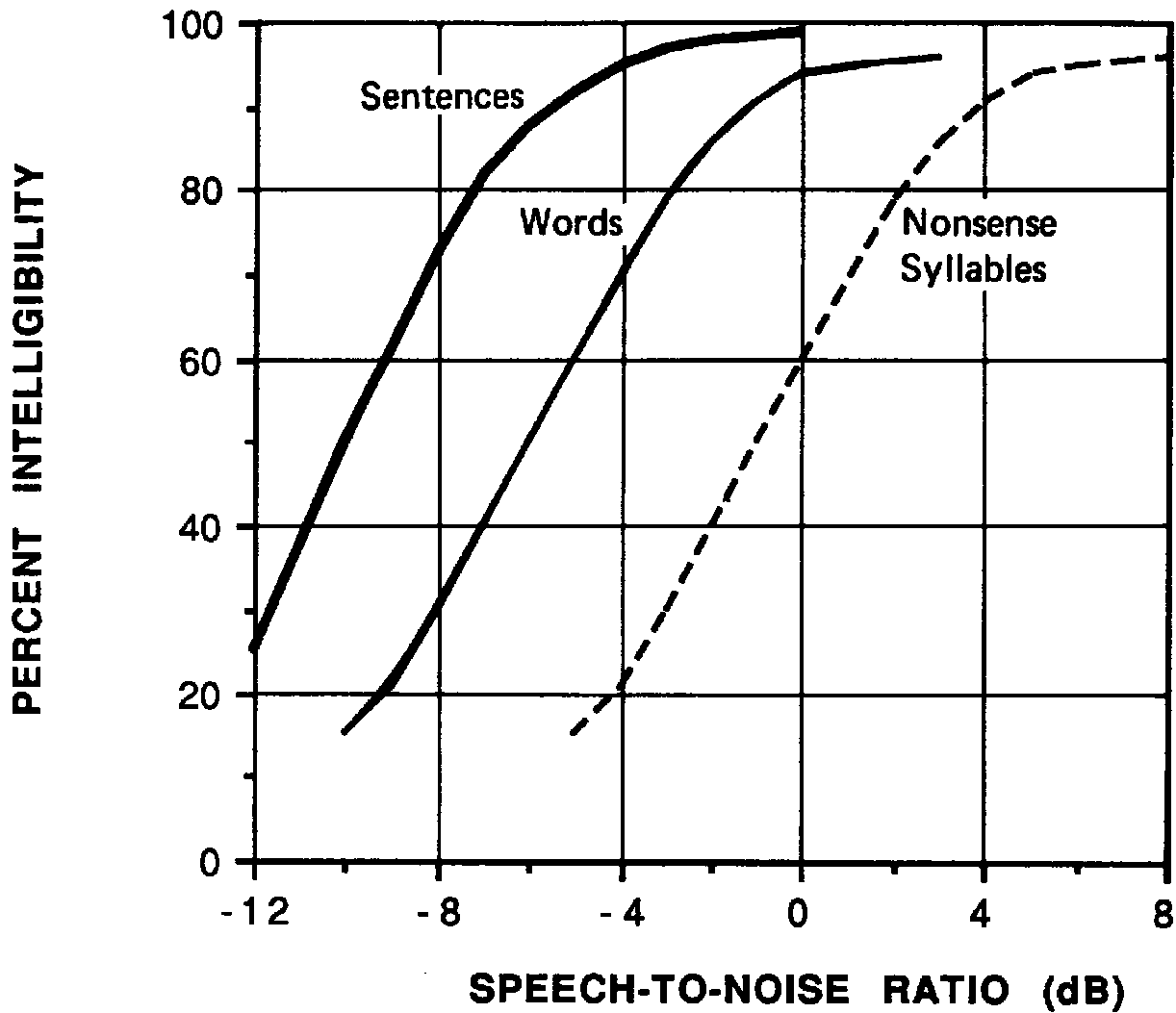


Figure 1. Speech intelligibility in noise for different types of test materials. Speech intelligibility is shown as function of speech-to-noise ratio for sentences, monosyllabic words, and nonsense syllables. These curves are approximate and depend on the test conditions, vocabulary size, and how the speech level is specified⁵.

⁵ Levitt, H. and Webster, J. C. Figure 16.3 of "Chapter 16. Effects of Noise and Reverberation on Speech," Handbook of Acoustical Measurements and Noise Control, edited by Cyril M. Harris.

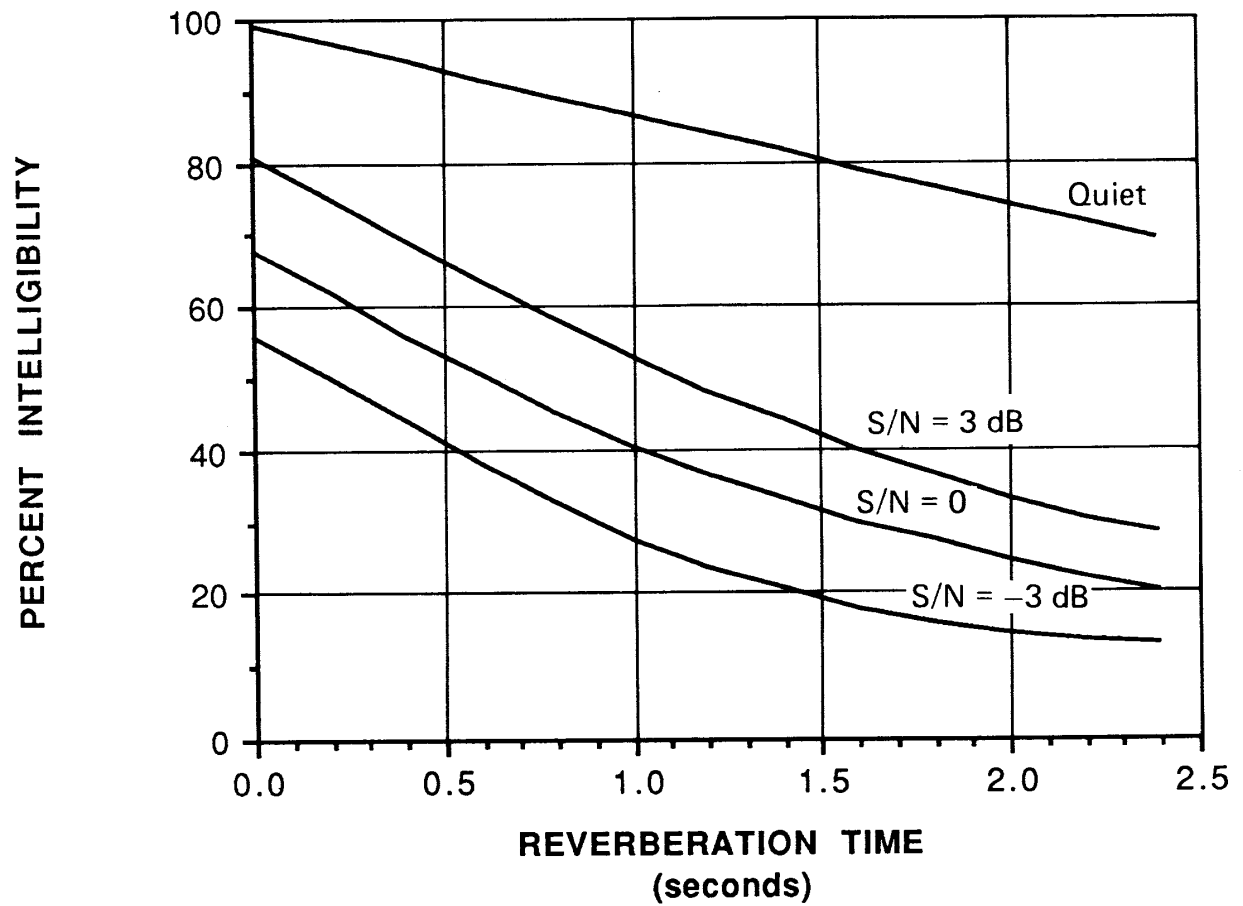


Figure 2. Speech intelligibility as a function of reverberation time and speech-to-noise ratio⁶.

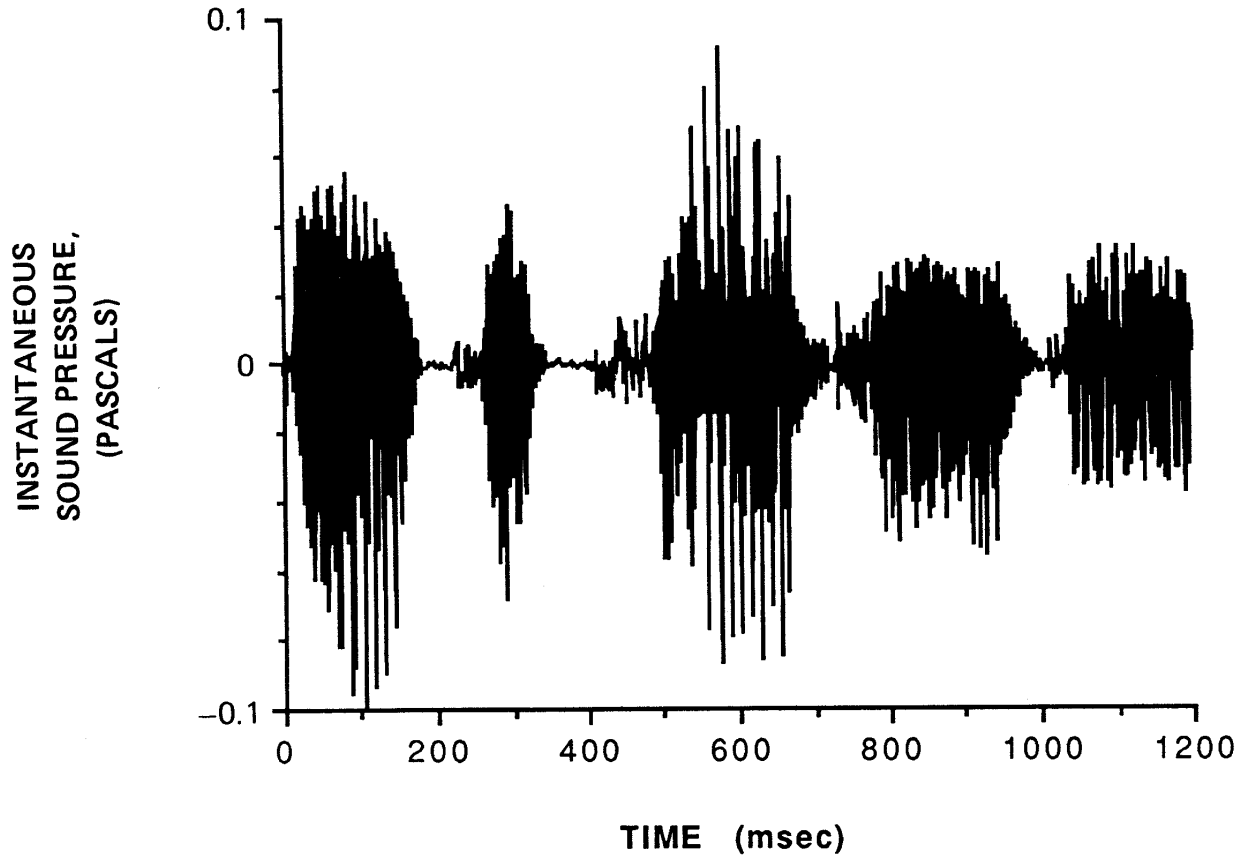


Figure 3. A typical speech waveform showing instantaneous sound pressure as a function of time. The resolution of the time scale is such that the waveform itself cannot be seen in any detail, but the amplitudes of the instantaneous pressure variations are clearly visible as a function of time at a distance of 1 meter⁷.

⁶ Levitt, H. and Webster, J. C. Figure 16.6 of "Chapter 16. Effects of Noise and Reverberation on Speech," Handbook of Acoustical Measurements and Noise Control, edited by Cyril M. Harris.

⁷ Levitt, H. and Webster, J. C. Figure 16.2 of "Chapter 16. Effects of Noise and Reverberation on Speech," Handbook of Acoustical Measurements and Noise Control, edited by Cyril M. Harris.