Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility

Dr. Herman Steeneken
Senior Research Scientist, TNO Human Factors

Jan Verhave
Research Scientist, TNO Human Factors

Steve McManus
Senior Engineer, Gold Line® Corporation

Kenneth Jacob
Chief Engineer, Bose® Professional Systems Division

Presented at Reproduced Sound 17
Stratford-on-Avon, November 16, 2001

Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility

Herman Steeneken[1], Jan Verhave[1], Steve McManus[2], Kenneth Jacob[3]

Abstract

International, European, and North American codes and standards now require sound systems used for emergency purposes to meet or exceed a minimum level of speech intelligibility. Intelligibility standards are also routinely being used for non-emergency systems. In both kinds of systems, however, verifying compliance is time consuming and requires the use of complex instrumentation and highly skilled technicians. As a result, the routine measurements of intelligibility necessary to enforce the codes and standards are rarely if ever made. With this problem in mind, the authors set out to develop a dedicated instrument that would accurately measure speech intelligibility according to international standards, and do so quickly and simply, thus making it possible for experts and non-experts alike to quickly measure intelligibility. Their results show clearly the feasibility of such an instrument.

[1] TNO Human Factors, Soesterberg, Netherlands
[2] Gold Line Connector Inc., West Redding, USA
[3] Bose Corporation, Framingham, USA

Introduction

Speech intelligibility is arguably the most important dimension of sound quality in most commercial and professional audio installations. Unfortunately, it is also the most common source of complaints from owners, operators, and the public. Why?

The answer can not be based on a lack of attention from the scientific, engineering, and regulatory communities. Speech intelligibility has received enormous attention from these communities for at least fifty years. For example, researchers have developed a number of proven methods for predicting intelligibility. Software programs for predicting the intelligibility of a design before it is installed – even before a building is constructed – have been developed and are used extensively. And technical codes and standards covering speech intelligibility are widely used by governments, facility owners, architects, audio consultants, and other professionals throughout the world.

In spite of this progress, measuring intelligibility using any of the proven methods remains a relatively difficult and expensive undertaking. As a consequence, routine surveys of existing facilities are rarely performed; sound system engineers design for good intelligibility but then rarely measure to verify performance; clients make intelligibility a high priority but typically do not have the final intelligibility level confirmed; and finally, because intelligibility measurements are so difficult and expensive to make, codes and standards writers are reluctant to require periodic compliance measurements even though they recognize that such measurements are essential to effective enforcement.

Fundamentally, then, achieving a desired level of speech intelligibility is essentially an open-loop process because feedback in the form of confirmatory measurements is so rare. As with any open-loop process, convergence can be slow and uneven, and is in fact not guaranteed. This suggests why, in the presence of almost universal agreement as to the importance of speech intelligibility, actual performance continues to generate a large number of complaints.

The problem is that speech intelligibility is not a simple parameter to measure. If it were, engineers would have developed accurate and easy-to-use instruments years ago. The proven methods of predicting speech intelligibility from physical quantities require relatively complex instrumentation and post-measurement signal processing. What has made the problem more tractable recently is 1) the availability of powerful, yet cost-effective microprocessors, and 2) research that shows how the computational demands of a proven measurement method can be reduced without sacrificing accuracy in the intended applications. Armed with both, a collaborative effort was undertaken to develop a fast, accurate, portable, and easy-to-use intelligibility prediction meter. The remainder of this paper describes the approach taken and the results obtained.

Speech Transmission Index (STI) considerations

A proven method of predicting intelligibility from physical quantities is the Speech Transmission Index, or STI method,[1,2] developed at TNO Human Factors in the Netherlands. Its accuracy, robustness, diagnostic capabilities, and wide applicability have made it a nearly unanimous choice by writers of codes and standards throughout the world.

In systems where linearity can be guaranteed, the STI can be measured relatively quickly by measuring the system's transfer function, and then using a mathematical formula to obtain the data necessary to compute the STI.[3,4] Unfortunately, linearity can not in general be guaranteed; in fact many systems have diminished intelligibility due for example to amplifier clipping and overdriven loudspeakers. Regardless, a test for linearity must be conducted first before the transfer-function method of calculating the STI can be used safely. Such a linearity test has complexities that would greatly compromise our goal of developing a simple yet accurate intelligibility prediction meter. Any technique that relies on measurement of a system's transfer function must therefore be rejected given that systems in which non-linearities are responsible for degradation of speech intelligibility are a fact of life, and given our goal of creating a fast and easy-to-use instrument.

To measure the STI of an arbitrary sound system, including systems where non-linearities play a role, a more direct approach must be taken. The STI method is based on the amplitude modulation of octave bands of noise, where the modulation frequencies and octave bands are chosen to match those of natural speech. Optimally, fourteen different modulation frequencies and seven octave bands are used, making a total of 98 different combinations of modulated noise. The STI method is based on research showing that the loss of modulation from a system's input to its output – which represents the loss of the modulations in natural speech – is also a measure of the loss of intelligibility. These so-called modulation reduction factors for the 98 combinations are clipped, weighted, and averaged to obtain the final STI value.

To measure the modulation reduction factor for just one of the 98 combinations requires that at least several periods of the lowest modulation frequency of interest be passed through the system. Other considerations such as error detection (discussed in more detail below) mean that a measurement that includes all 98 combinations would take on the order of 10-15 minutes to complete. Such a measurement, which would yield a value for only one position in a room or one condition in a sound system, misses the target for a fast measurement by at least an order of magnitude, and must also therefore be rejected as an option given our goals.

Measurement time considerations are not new to the researchers and users of the STI method. In the early nineteen eighties, research was conducted at TNO to determine if a smaller set of the 98 combinations of amplitude-modulated noise bands could be used in certain applications with an acceptable loss of accuracy. A subset of nine combinations, called RASTI (for Room Acoustics Speech Transmission Index or Rapid Speech Transmission Index) was shown to be effective in some applications, and a commercial instrument was produced embodying the technique.[5,6] Only about eight seconds were required to make a RASTI measurement. However, the instrument was considered as a screening device for person-to-person communications because the excitation signal was limited to just two octave bands. This made the system unsuitable for measurements on sound systems in rooms, which are known to vary widely in their performance from octave band to octave band.

The explosive advancements in micro-circuitry and software in the years since RASTI was developed compelled these authors to revisit the question of making fast and accurate STI measurements suitable for use in the professional and commercial audio field. First, the team felt it was possible to drastically reduce the bulk of equipment required on the reception-and-processing end of the measurement by using digital micro-circuitry; a bulky piece of equipment requiring a power cord could conceivably be reduced to a lightweight, handheld, battery-operable device. Second, the team felt a new subset of the 98 combinations of modulated noise could be designed that was optimized for public address systems.

STI-PA: an efficient form of the Speech Transmission Index method for public address systems

Given these considerations, research was undertaken to determine if a new subset of modulated noise bands could be shown to be an acceptable substitute for the full set of 98 modulated bands over a set of conditions designed to represent commercial and professional sound systems.

As an initial starting point, the team chose to include all seven octave bands rather than the two used in RASTI measurements. Eliminating whole octave bands at the outset was avoided because professional and commercial sound systems can suffer from degradation in intelligibility in any of the relevant frequency bands. However, it is well known that the lower octave bands of speech play a much smaller role in speech intelligibility than the middle and higher bands.[7] Thus a portion of the research focused on the question of whether separate 125 and 250 Hz octave bands were justified. Additionally, the team considered simultaneous modulation by two or more frequencies in each octave band as an option. Tempering this potential advantage, however, was the knowledge that addition of a second, third or greater number of modulation frequencies in a single octave band would have the effect of reducing the dynamic range of the excitation signal and thus reducing the accuracy of measurements in situations where the ambient noise level would be very low.

In the end, the team arrived at a six-band, two-modulation-frequencies-per-band signal. The six bands consist of the 125 and 250 Hz octave bands combined, together with the 500 Hz - 8 kHz octave bands. The STI value obtained using the new subset was compared to the STI value obtained using all 98 combinations in a large number of conditions intended to represent systems in which intelligibility is degraded by reverberation, noise, and band limiting. The results, showing more than 800 conditions, are shown in Figure 1. The agreement is outstanding, proving the value of the new subset as a computationally efficient and viable substitute for the full 98 combinations of modulated noise.

The name “STI-PA” (for Speech Transmission Index – Public Address) was chosen to differentiate this subset of modulated bands from others already in use (e.g. RASTI, or STITEL, which is designed for testing telephone systems).

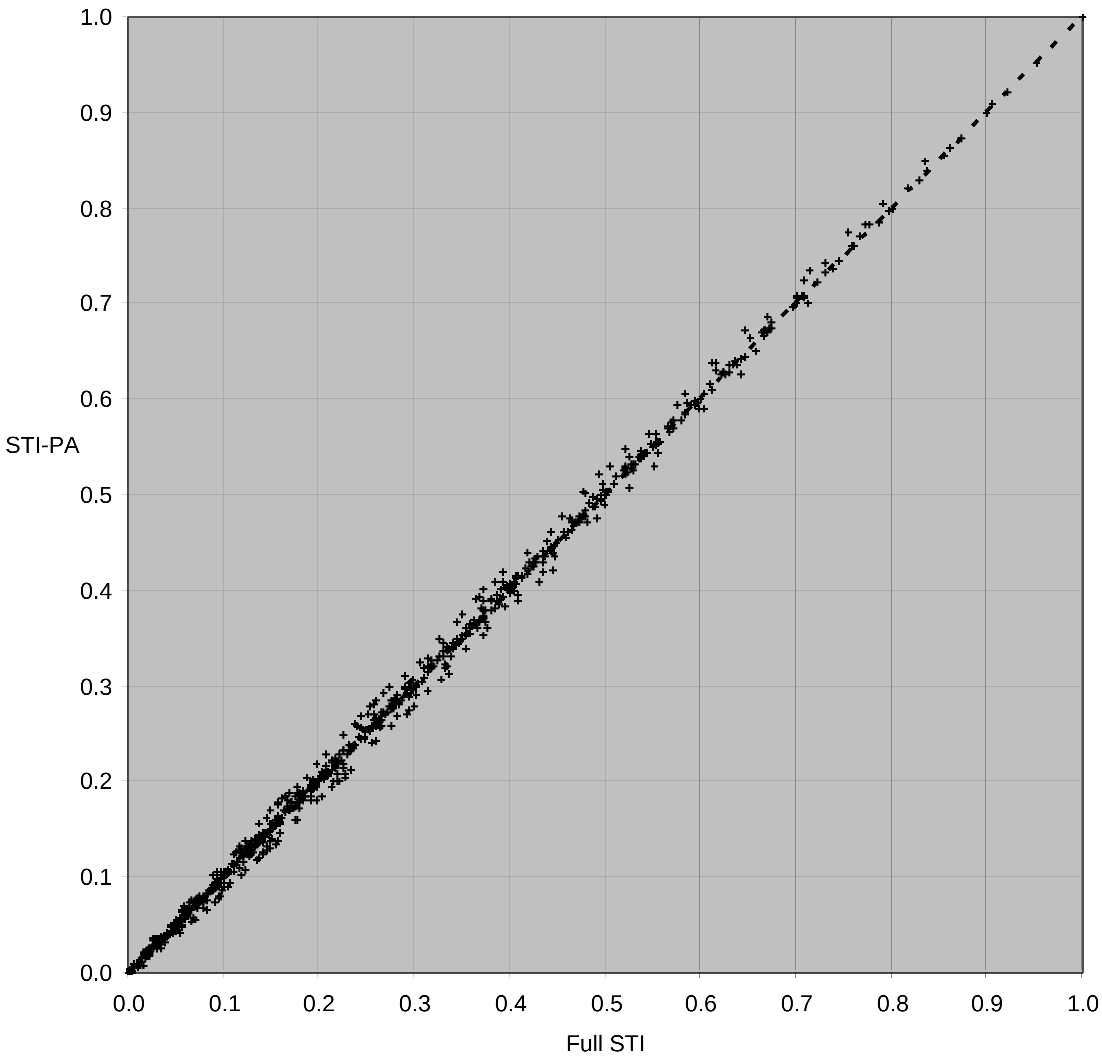

Figure 1. STI values calculated using the new STI-PA excitation signal are compared to STI values calculated using the full 98 combinations of modulated noise. The 45° line represents perfect agreement. A total of 880 different conditions varying reverberation times (0.00, 0.25, 0.50, 1.00 2.00, 4.00, and 8.00 seconds), signal-to-noise ratios (-15, -12, -9, -6, -3, 0, 3, 6, 9, 12, and 15 dB), and bandpass filters (2 kHz, 1-2 kHz, 1-4 kHz, 1-8 kHz, 500 Hz - 2 kHz, 500 Hz - 4 kHz, 500 Hz - 8 kHz, 125 Hz - 2 kHz, 125 Hz - 4 kHz, and 125 Hz – 8 kHz) are shown. In 833 out of 880 conditions (95%) the error is less than ± 0.02. The maximum error is ± 0.03. The agreement is outstanding, proving the suitability of the STI-PA signal in public address applications.

Embedded system vs. general purpose computer

With the results showing that a new, computationally efficient form of the STI could be used, attention turned to the issue of implementing the reception and signal processing functions on a micro-processor-based measurement device. While the necessary processing certainly could be made to fit on a general-purpose computer, an embedded system was selected in order to guarantee an instrument that was simple to use and reasonable to calibrate and service.

From one perspective, this decision may seem unintuitive since the cost of a powerful laptop computer is now comparable to an embedded-system device, and can do so much more. And yet that is exactly why the general-purpose computer was rejected. The fact that it can ‘do so much more’ is the very reason that each user-customized unit tends to behave differently, and why one machine’s results often fail to match another’s, both anathema to the fundamental principle that measurements should be repeatable from one like machine to another.

Moreover, the instrument envisioned is meant to be used in code-compliance measurements (among other things) which means that proper service, calibration, and maintenance will be essential if owners, insurance companies, government officials, and others are to completely trust the results. The idea of providing the required level of service on a multitude of general-purpose computers, each of which had been customized by the user, was rejected as unmanageable.

User interface

TO reach the goal of making a meter that would be truly easy to use, we envisioned a user interface consisting of a single button that would initiate a measurement and a counter that would inform the user how much time remained to complete the measurement. With the excitation signal running continuously, the user would initiate a test, and about fifteen seconds later would get an intelligibility score based on the STI.

One of the most important and widely applied standards, *IEC 60849: Sound Systems for Emergency Purposes*, allows the use of a number of different methods for measuring speech intelligibility, including the STI. In order to relate the results of the different methods, a Common Intelligibility Scale (CIS) was created. The actual language in IEC 60849 calls for a minimum CIS value, which is therefore what we also chose to display. The meter computes the STI and then converts it to the CIS.

Error checking

Because the anticipated users of the meter include non-experts in acoustics (inspectors, local authorities, etc.) as well as acoustical experts, the team felt it was important that the meter have robust error detection capabilities. In the case of an intelligibility prediction meter upon whose output could depend the granting of an occupancy permit, the outcome of a lawsuit, or the rightful collection of professional fees, the desire to notify the user when accurate measurements could *not* be made was considered of paramount importance.

To address this need, error-checking algorithms were developed and implemented in the meter. For example, a common occurrence when making acoustical measurements of any kind, including STI measurements, is for conditions to be interrupted by spurious noise. If someone talks next to the microphone during a measurement, or if a worker drops a piece of construction material, for example, the data generated are suspect. Rather than rely on the judgement of the user in such situations, the meter automatically invalidates measurements when the interfering noise produces an unacceptable error. In most cases, the test can then simply be re-run.

Instrument testing

Prototypes of the meter[8] were tested to ensure that they behaved according to theory and met our goals for accuracy, portability, and simplicity. STI values obtained on the meter were compared to STI values obtained using the full 98 combinations of modulated noise as calculated on a reference system at TNO. A number of representative test conditions were used, including different bandpass conditions, noise levels, non-linear effects, reverberation and echo profiles. The results are shown in Figure 2. The data show conclusively that the prototype meter is accurate in a wide range of conditions typical of public address systems.

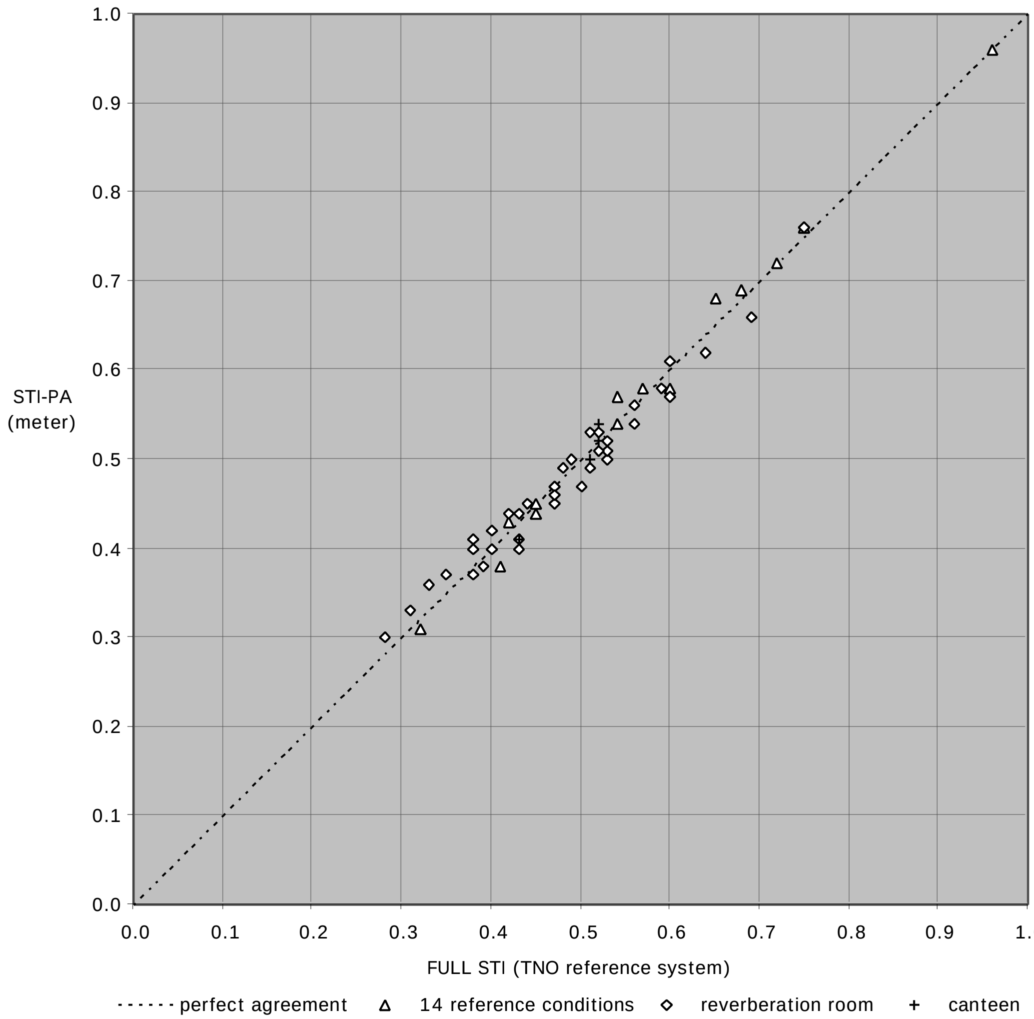

Figure 2. For a variety of conditions, STI values from the prototype meter are compared to values obtained on an STI reference system at TNO that uses the full 98 combinations of modulated noise. The test conditions are: 1) Fourteen reference conditions which use a mixture of band pass limiting, noise, peak clipping, and echoes.[9] 2) Measurements made in a TNO reverberation room using different numbers of absorbers in order to vary the reverberation time (0, 1, 2, 4, 8, and 16 absorbers resulting in a reverberation time range of 0.2 – 4.0 seconds), different bandwidths (500 Hz - 4 kHz and 1-4kHz) and different noise conditions (no noise and 50 dBA of speech babble), all using an artificial talker as a source producing 50 dBA at one meter. 3) Measurements from the TNO canteen, with either a PA system or an artificial talker as a source, and at two locations (5 and 15 meters from the artificial talker). Most data are within ± 0.02 and the worst case error is only ± 0.03. The results show that the meter is very accurate over a broad range of conditions common in public address systems.

Conclusion

The central assumption here is that if the world wants better intelligibility – which it apparently does based on the proliferation of codes and standards requiring it – then we must have a simple and effective means to measure it. Without such means, intelligibility will continue to be of intense scientific and technical interest but the public will not benefit in any fundamental or widespread way.

With these concerns in mind, these authors felt that it might be possible to develop a speech intelligibility prediction meter that worked almost as easily as a sound level meter. The approach and results presented here are intended to build the confidence of other experts that such a meter is feasible and has been successfully demonstrated.

Such a device may trigger concern among the community of experts who currently make intelligibility measurements. If others who are not necessarily acoustical experts can now make intelligibility measurements, does this threaten the expert's income? We believe the opposite is true – that the existence of a simple and effective way to measure intelligibility will greatly *increase* the demand for expertise, not jeopardize it. Who will be called on to diagnose and recommend changes to a system that fails an intelligibility test? Similarly, who will be called on to create a design that is certain to pass the test in the first place? Frankly, we suggest the problem may be that we lack *enough* experts to meet the demand that will be created.

Fundamentally, our hope is that the single-most important dimension of sound quality – speech intelligibility – undergoes significant improvement in the coming years. Spoken announcements remain one of the most effective and vital means of communication in public places and places of business – for both emergency and non-emergency purposes – but only if they are intelligible. Together with the scientific knowledge of how to measure intelligibility, and many of the codes and standards that allow the public and private sectors to require it, we submit the 'third leg' now exists: a simple and effective means of measuring it.

[1] T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.*, vol. 77, No. 3 (1985).

[2] H. J. M. Steeneken and T. Houtgast, "Mutual Dependence of the Octave-Band Weights in Predicting Speech Intelligibility," *Speech Commun.*, vol. 28, pp. 109-123 (1999 June).

[3] M. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *Acustica*, vol. 49 (1981).

[4] H. J. M. Steeneken and T. Houtgast, "A Physical Method for measuring Speech Transmission Quality", *J. Acoust. Soc. Am.*, vol. 67, 318-326 (1980).

[5] T. Houtgast and H. J. M. Steeneken, "RASTI: A Tool for Evaluating Auditoria", B&K Technical Review (1985).

[6] B&K Speech Transmission Meter 4225, and Receiver 4419.

[7] T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Am.*, vol. 77, No. 3 (1985).

[8] Gold Line DSP30 with STI-CIS™ module.

[9] The fourteen conditions used are described in H. J. M. Steeneken, "On Measuring and Predicting Speech Intelligibility," Ph.D. thesis, University of Amsterdam, pp. 143-144, (1992). The abbreviations for the fourteen conditions used here are:
BP01 - BP1 + SNR inf; BP04 - BP1 + N2 + SNR = 0; BP08 - BP1 + N4 + SNR = 3;
BP17 - BP2 + N4 + SNR = 6; NL01 - PC1 + BP1 + SNR = inf; NL04 - PC1 + BP1 + N4 + SNR = 9;
NL08 - PC1 + BP2 + N1 + NSR = 6; NL14 - CC2 + BP1 + N4 + SNR = inf;
E01 - E1 + BP1 + SNR = inf; E04 - E2 + BP1 + SNR = inf; E06 - E2 + BP1 + N4 + SNR = 6;
E07 - E3 + BP1 + SNR = inf; E08 - E3 + BP1 + N4 + SNR = 12; E09 - E3 + BP1 + N4 + SNR = 6.