Technical Note

Note: Technical notes are aimed at promoting discussion. The views expressed are not necessarily those of the editors or the Australian Acoustical Society. Contributions are not formally peer-reviewed.

THE SPEECH TRANSMISSION INDEX AFTER FOUR DECADES OF DEVELOPMENT[1]

Sander van Wijngaarden, Jan Verhave and Herman Steeneken
Embedded Acoustics BV, Delft, The Netherlands
sander@embeddedacoustics.com

This year, the Speech Transmission Index (STI) celebrates its 40th birthday. It has been four decades since Houtgast and Steeneken first published their objective method for predicting speech intelligibility in Acustica [1]. Since then, the STI has evolved into a versatile and mature method, used in a diversity of applications. It is now more popular than ever, with record numbers of STI users as well as manufacturers of STI measuring solutions. We mark the occasion by looking back at the development of the Speech Transmission Index throughout the decades, while also presenting an overview of current developments and challenges.

ORIGINS OF THE SPEECH TRANSMISSION INDEX

What inspired Houtgast and Steeneken to develop the STI was their desire to save time and to eliminate the dull work associated with subjective intelligibility tests. Or, in the words of Houtgast: their "laziness". Their work back then, at TNO in the Netherlands, consisted largely of carrying out lengthy evaluations of speech intelligibility, mainly of military communication systems, using large numbers of human test subjects. The need was there for a faster, and more diagnostic, alternative to subjective listening tests. The primary design objective was that it should be a physical measuring method (ie. based purely on physical principles without humans in the measuring loop), which could produce results fast. Moreover, a measuring method was required that could use a test signal in order to obtain direct and immediate results. This sets the Speech Transmission Index apart from the Articulation Index (AI), which was already around at the time. The STI owes several of its key characteristics to the work done by French and Steinberg [2] on which the AI is also based. However, the AI (and later on its successor the Speech Intelligibility Index, SII) is basically calculated from measured sound pressure levels, theoretical data or measured impulse responses. Among other things, this means that the AI and SII are inherently "blind" to non-linear effects, whereas the STI incorporates these effects.

The Speech Transmission Index concept also incorporated insights crossed over from research in the visual domain in the early 1970s. Optical system engineers back then already used the concept of the Optical Transfer Function (more generally named the Modulation Transfer Function) to quantify the transmission quality of optical systems. Houtgast and Steeneken realized that similar principles in the time domain should apply to transmission of speech signals.

[1] Originally published in the IOA Acoustics Bulletin, May/June 2012 and reprinted with permission.

KEY CONCEPT

Houtgast and Steeneken designed their STI test signals based on modulated, speech-shaped noise. The basic principle underlying the STI is that preservation of speech intelligibility during transmission is achieved by preservation of the natural intensity fluctuations in speech spectra. The design of test signals was such that they mimicked these natural modulations, but in such a way that measurements could be carried out quickly, precisely and within the constraints of calculation (computer) power of the time. After four decades of evolution, the basic principles remain unchanged – although the computer power is now available in handheld devices, whereas the necessary equipment originally required several people to lift.

INITIAL USE OF THE STI METHOD

In the 1970s, the STI was very much a niche method. The inventors themselves used the STI in various real-life applications, but use by others was limited to a few studies done out of scientific interest only. The publication of Steeneken and Houtgast's JASA paper in 1980 [3] marked the beginning of more widespread use of the method. The growing group of STI users forked into two separate (but overlapping) communities almost from the very beginning. On the one hand, there is a scientific community, attracted to the way the STI predicts speech intelligibility based on a near-universally applicable model with only few design parameters. On the other hand, there is the engineering community, interested mostly in the practical advantages that the STI was designed for: fast, objective and accurate predictions of speech intelligibility.

To the engineering community, standardization of the STI method by successive IEC-committees (in successive editions of IEC 60268-16 [4]) turned out to be of key importance. The version of the STI described by Steeneken and Houtgast [3] was standardized as the original, first edition of IEC 60268-16. TNO already had a variety of test signals available, but the

RASTI test signal (Room Acoustical STI), designed specifically for application of the STI in room acoustics) saw the most widespread use. This was largely due to the availability of RASTI measuring hardware from HYPERLINK "http://www.bksv.com/" Brüel & Kjær, based on TNO's earlier RASTI device (figure 1).

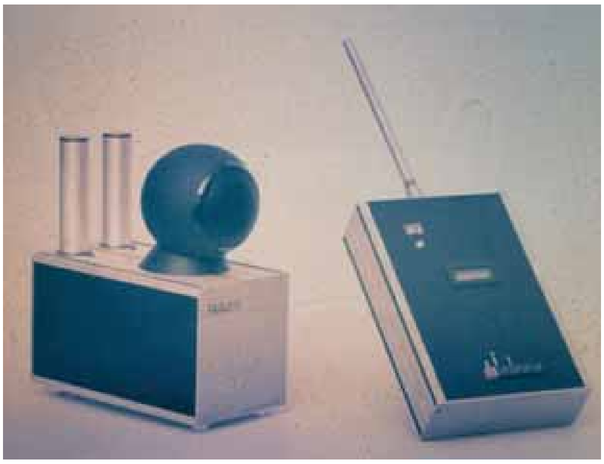

Figure 1. First handheld implementation of the STI (RASTI hardware, 1980)

Over the years, a lot of criticism towards the STI came from users having experiences with RASTI outside its intended scope of use. RASTI measurements are accurate measurements of the STI, if applied to pure room acoustics; ie. transmission chains featuring electro-acoustic components should never be measured using RASTI. Words to this effect in the RASTI manual have not stopped people from attempting to do so anyway – and even publishing criticizing accounts of how RASTI failed to yield accurate predictions.

IEC 60268-16 SECOND EDITION (1998)

There was also a certain amount of justified criticism towards the "original" STI, which triggered a significant amount of research at TNO in the 1980s en 1990s to improve on the method. Several major improvements were standardized in the second edition of IEC 60268-16, which was released in 1998. The original STI did not account for the fact that speech perception is aided by synergistic effects between adjacent frequency bands. Among several other improvements, additional model parameters were added to take these between-band interactions into account. The 2nd edition of the STI was named STI_r ('r' for revised), but the subscript was dropped later on. It is now customary to simply refer to any version as "STI," indicating which revision of the IEC standard applies in accompanying text (if relevant).

The STIDAS IID device produced by TNO was capable of measuring the STI according to first and second editions, using a host of different test signals, including full STI modulated noise test signals and STITEL (specifically for telecommunication measurements). This device was sold worldwide, but its specific hybrid analog-digital design made it too expensive for many users. Some of these units remain in service to date, mostly at military research facilities. A photo of the STIDAS STI device is shown in figure 2.

Figure 2. STIDAS I (STI Device using Artificial Signals) device based on a PDP-11/10 computer and custom analog hardware (1971)

A trend in the 1990s was that many acousticians started to use estimations of the STI based on measured impulse responses. Affordable PC-based software for impulse response measurements was becoming commonplace. If certain conditions are met (among which linearity, no back ground noise or band-pass limiting), then the STI may be precisely derived from the impulse response. This is what many users were doing (or rather, what their software was doing for them). Unfortunately, the conditions for this approach to work do not generally apply. In fact, much like RASTI, impulse response-based STI estimates can only be relied upon in evaluations concerned purely with room acoustics. A need was widely felt for a test signal (and a version of the STI method) that was applicable to electro-acoustics transmission chains, and could be measured quickly and directly. This led to the development of STIPA.

IEC 60268-16 THIRD EDITION (2003)

The third edition introduced two major changes. Most importantly, it introduced the STIPA test signal (sometimes referred to as STI-PA), which is a test signal optimized for PA systems. Compared to RASTI, STIPA has the advantage that all octave bands are covered (125 Hz – 8 kHz), although only two modulations frequencies are tested per octave band. This means that STIPA can be used reliably in nearly all cases involving electro-acoustics as well as room acoustics. STIPA can be used in any condition that RASTI was previously intended for, with the possible exception of rooms featuring pronounced, individual echoes. Since RASTI is inherently unsuitable for any condition involving electro-acoustics, the introduction of STIPA made RASTI completely obsolete.

The 3rd edition also introduced the concept of level-dependent masking. Earlier versions of the STI ignored the fact that auditory masking curves flatten out at higher sound levels, effectively reducing intelligibility. The resulting mismatch sometimes observed between the STI and subjective intelligibility at high sound levels no longer exists from the 3rd edition onwards. The price for this added accuracy is that measurements need to be calibrated in terms of the (A-weighted) sound pressure level. This was already common practice, but not specifically required before. If acoustic calibration is not feasible (e.g., when evaluation intelligibility of purely electronic devices that may be used at arbitrary speech levels), level dependent masking may be disabled. The resulting STI is then only valid for comfortable listening levels.

The design and release of STIPA had the intended effect. A photo of the first STIPA-capable device to reach the market is shown in figure 3. Measuring devices by several manufacturers reached the market, and the last users that had been holding on to their now-obsolete RASTI equipment made the transition. Although STIPA is just one of several standardized test signals in the 3rd edition, it turned out to be virtually the only one used in practice. Many users still using indirect (impulse-response based) measurements also decided to obtain STIPA-capable devices. Some (local) regulations specifically requiring STIPA helped to speed up this process. In practice, situations for which the STIPA test signal is insufficient, and "full STI" measurements are required, are rare; this is the case mainly when strong discrete, single echoes occur.

Figure 3. The first STIPA-capable device to reach the market, made by Gold Line (2002)

IEC 60268-16 FOURTH EDITION (2011)

Even if the STI method itself had some room left for future improvement in its third edition, it was mostly the text of the IEC standard itself that now became criticized. With more equipment manufacturers implementing STIPA, it became apparent that it was not easy to build a STIPA-capable device when using the standard as a single source of information. The standard was therefore completely overhauled and much information was added.

The fourth edition of the standard [4] outlines not only how to design direct STI measurement (using modulated test signals such as STIPA) but also how to implement indirect (impulse response-based) measurements. Limitations of different approaches and test signals are now clearly indicated in the standard. In other words, for different types of application, the standard now prescribes which methods may, and which ones may not be used safely.

The fourth edition features only a single (minor) change to the STI algorithms itself: the calculation of level-dependent masking was changed from a discrete lookup-table to a smooth continuous function. Also added is information on interpretation of the STI relative to true speech intelligibility. Whereas the STI quantifies the impact of the transmission channel on intelligibility, there is also an influence of talkers and listeners. There are fixed and well-known relations between STI and intelligibility for "normal" populations. The 4th edition of the standard also assists in interpreting the STI for populations of non-native talkers and listeners, as well as certain categories of listeners with hearing loss.

THE MAJOR CURRENT CHALLENGE: VALIDATION AND CERTIFICATION

Every successive update of the STI method was validated at TNO, using a reference system called COMCHA. This reference system simulated a wide variety of representative test channels (78 channels based on band-pass limiting and 68 channels for communication channels). TNO also maintained reference versions successive generations of measuring devices. Besides validation of new additions to the STI framework, these tools were also used to provide third-party validation and certification services, for instance for STIPA measuring devices from various manufacturers.

Today, validation services based on these assets are no longer be offered by TNO. In practice, there is no other institute or company capable and willing to take over this service that has the same level of confidence, expertise and (especially) independence. This is perhaps the major current challenge for the future of the STI: making sure that all STI devices measure consistently and correctly according to the standard and produce identical results. Likewise, all STIPA signals (and also other STI test signals), should be interchangeable and compatible with each IEC-compliant measuring device.

For the moment, the best solution appears to be to create an open-source validation database. TNO and Embedded are collaborating in creating such a reference database of degraded STIPA test signals using the original COMCHA conditions, verified with "golden standard" software from TNO. This set of signals will represent the various types of conditions for which STIPA is sensitive, such as noise, reverberation, peak clipping, etc. This database will be made available through the internet under an open licensing regime, such as (for instance) GPL. Not only will developers be able to test and validate their devices using these signals; their users (and competitors) will be able to check compliance using the very same database. In our view, this provides for a system of checks and balances that eliminates the need for an impartial certifying authority.

CURRENT AND FUTURE RESEARCH

The STI has been a tool in many scientific studies, but it is also itself the subject of scientific investigation. In the past, the focus was often to improve the method, in terms of solving known inaccuracies and issues with the method. Nearly all of these issues have been thoroughly investigated and are now closed chapters; examples are the interaction with gender, non-linear auditory masking and variations in the modulation spectrum. Right now, the focus of research is more on *extending* the scope of the method rather than just generally improving it.

One very interesting field of research is measuring the STI using real, recorded, speech instead of artificial test signals. This was actually considered from the very beginning; in the early years however, there was simply a lack of processing power for this to be practically feasible. First accounts of speech-based STI measurements were published in the 1980s. A difficulty with speech-based STI measurements is that useful, natural modulations are present (such as in the artificial test signals), but detrimental components, such as nonlinear distortion components, tend to have similar modulation spectra. Alternative approaches were proposed, among others, by Drullman [5] and Payton [6], but their approaches were only partially successful in separating between useful and detrimental modulations. The concept of weighing modulations frequencies within an MTF based on the question whether or not phase shifts occur was explored by van Gils and van Wijngaarden [7], and proven promising. Speech-based STI measurements were, among other applications, shown useful to evaluate digital voice coders. An open question at the moment is to decide on optimal phase weighting functions. Also, further validation in a wider range of realistic conditions is needed.

Another field of research is the study of binaural STI measuring methods. The STI has always been a monaural model. This means that the STI cannot be used to distinguish between conditions in which binaural listening benefits are significant. Specific model additions have been proposed by van Wijngaarden and Drullman [8] to incorporate binaural listening. Similar work has been done by Beutelmann et al. [9] in the context of the Speech Intelligibility Index (the successor to the Articulation Index). This work needs to be consolidated into a robust addition to the STI model, that may optionally be used to refine STI-based studies in which binaural listening plays a predominant role. Such an addition also needs to be validated.

MEASURING THE "FULL" STI WITH MODULATED NOISE CARRIERS

Another relevant current research topic is concerned with improving and extending the current array of test signals. At the moment, the STIPA test signal is used nearly exclusively. This means that only two modulation frequencies per octave band are tested. A "full" STI measurement involves modulation frequencies sampled in 1/3 octave bands from 0.63 Hz to 12.5 Hz. In practice, a sparsely sampled MTF matrix (such as the one offered by STIPA) suffices for most applications – but not all. As mentioned above, care should be taken when using STIPA in rooms with discrete echoes. All current commercially available methods for measuring the full MTF matrix make use of inverse calculation of the MTF based on impulse response measurements. This is not permitted if nonlinear distortion components may occur. Only the TNO reference system currently features a fully IEC-compliant measurement mode for full STI measurements. The drawback of the TNO system is that it is based on obsolete hardware, takes up to 10 minutes for a single measurement point, and requires the test signal generator and the STI analyzer to be synchronized.

Embedded Acoustics has initiated a research project that is intended to result in an advanced full STI measuring scheme, based on modulated noise carriers, that does not need to be synchronized. In practice, a measurement will appear to be similar to a STIPA measurement, except for the measurement time (which will probably need to be 1 to 2 minutes).

ON TO THE NEXT FOUR DECADES…

When the 4th edition of IEC-60268-16 was published last year, hardware and software vendors proved quick to update their products. This is encouraging; it shows that the market is quick to respond to changes. Several companies will launch new STI products in 2012, from STIPA modules for existing hardware to completely new devices and mobile apps (figure 4). Also, the STI is finding its way into new standards and regulations every year, replacing now-obsolete subjective intelligibility tests and less advanced metrics. This ranges from the national NEN-2575 standard for certification of Voice Evacuation systems in the Netherlands, to the NFPA-1981 standard in the US for testing speech intelligibility of face masks.

In conclusion, there is a community willing and able to support the STI, and the number of users is also consistently growing. Keeping the method up to date for another forty years will be an effort that requires this community of individuals and companies to actively cooperate. We predict that in the next few years we will see this community pulling together, and starting to prepare work for the 5th update of the IEC standard, somewhere around 2016.

Figure 4. iPhone apparatus for performing 4th edition-compliant STIPA measurements (2011)

REFERENCES

[1] T. Houtgast and H.J.M. Steeneken, “The modulation transfer function in room acoustics as a predictor of speech intelligibility”, *Acustica* **28**, 66-73 (1973)

[2] N.R. French and J.C. Steinberg, “Factors governing the intelligibility of speech sounds”, *Journal of the Acoustical Society of America* **19**, 90-119 (1947)

[3] H.J.M. Steeneken and T. Houtgast, “A physical method for measuring speech-transmission quality”, *Journal of the Acoustical Society of America* **67**, 318-326 (1980)

[4] International Electrotechnical Commission IEC 60268-16 *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*, Fourth edition, 2011

[5] R. Drullman, J.M. Festen and R. Plomp, “Effect of reducing slow temporal modulations on speech reception”, *Journal of the Acoustical Society of America* **95**, 2670-2680 (1994)

[6] K.L. Payton and L.D. Braida, “A method to determine the Speech Transmission Index from speech waveforms”, *Journal of the Acoustical Society of America* **106**, 3637-3648 (1999)

[7] B.J.C.M. van Gils and S.J. van Wijngaarden, “Objective measurement of the speech transmission quality of vocoders by means of the Speech Transmission Index”, *Proceedings of the RTO Human Factors and Medicine Panel (HFM) Symposium: New Directions for Improving Audio Effectiveness*, Amersfoort, The Netherlands, 11-13 April, 2005

[8] S.J. van Wijngaarden and R. Drullman, “Binaural intelligibility prediction based on the speech transmission index”, *Journal of the Acoustical Society of America* **123**, 4514-4523 (2008)

[9] R. Beutelmann, T. Brand and B. Kollmeier, “Revision, extension, and evaluation of a binaural speech intelligibility model”, *Journal of the Acoustical Society of America* **127**, 2479-2497 (2010)