THE FUTURE IS BRIGHT FOR THE SPEECH TRANSMISSION INDEX; DEALING WITH NEW CHALLENGES AFTER FOUR DECADES OF DEVELOPMENT

SJ van Wijngaarden Embedded Acoustics, Delft, The Netherlands
HJM Steeneken Embedded Acoustics; former TNO, Soesterberg, The Netherlands
JA Verhave Embedded Acoustics, Delft, The Netherlands

1 INTRODUCTION

We are approaching the 40th anniversary of the first international publication on the Speech Transmission Index [1]. Time has not been standing still – nor has development of the STI method. For speech transmission channel designers trying to achieve maximum speech intelligibility, most challenges remain the same: dealing with reverberation, ambient noise, bandwidth, overload and loudspeaker distortion. Other things have changed; we generally do not have to worry about centre clipping of carbon microphones anymore, but we do have to be concerned about the effects of digital voice coding and noise suppression.

The STI has been doing remarkably well in keeping up with new challenges, largely due to a great number research projects carried out at TNO. At this point in time, it has become clear that TNO will not be supporting further STI-related research. Others have to step up and contribute, to keep the STI in shape for another 4 decades or so. Given the large base of knowledgeable STI users, there is sufficient potential, and also willingness to invest. The key is organisation: how do we get the STI community to join forces, and make further development of the STI a truly collaborative effort?

This paper does not only outline the history of the STI up till now, but also suggests a way forward. We will argue why we believe the future is bright for the STI.

2 THE STI IN HISTORICAL PERSPECTIVE

2.1 Evolution of method and algorithms

2.1.1 Origins of the STI method

The Speech Transmission Index was conceived by Tammo Houtgast and Herman Steeneken. The first international publication dates from 1971 and appeared in Acustica [1]. What inspired Houtgast and Steeneken to develop the STI was the desire to save time and to eliminate the dull work associated with subjective intelligibility tests. Or, in the words of Houtgast: their “laziness.” Work at TNO back then 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 (i.e., 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 [2]. The AI was already there at the time; the STI owes several of its key characteristics to the work done by French and Steinberg [3] on which the AI is also based. However, the AI (and later on its successor 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.

Figure 1. STIDAS I (STI Device using Artificial Signals) device based on a PDP-11/10 computer and custom analog hardware (1971).

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.

2.1.2 Key design characteristics

Houtgast and Steeneken designed 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 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.

2.2 Evolution of methodology, measuring devices and test signals

2.2.1 First widely used version (1980); first edition of IEC 60268-16.

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) turned out to be of key importance. The version of the STI described in Steeneken and Houtgast's 1980 JASA paper 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 B&K, based on TNO's earlier RASTI device (figure 2).

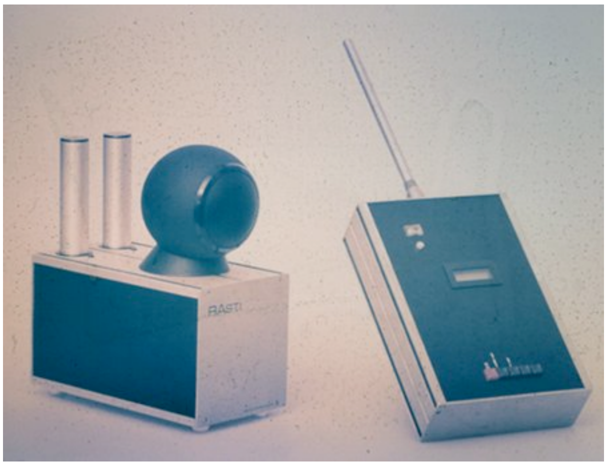

Figure 2. First implementation of RASTI in hardware (1980).

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; i.e., 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.

2.2.2 IEC 60268-16 2nd edition (1998)

There was also a certain amount of justified criticism towards the "original" STI, which triggered a significant amount of research at TNO [4 - 7] in the 1980s en 1990s to improve on the method. Several major improvements were standardized in the 2nd 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 STIr ('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.
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 [8].

2.2.3 IEC 60268-16 3rd edition (2003)

The 3rd edition introduced two major changes: introduction of the STIPA test signal (two modulation frequencies per octave band, 7 octave bands) and the introduction of level-dependent masking.

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

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. Measuring devices by several manufactures 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.

2.2.4 IEC 60268-16 4th edition (2011)

Even if the STI method itself had some room left for future improvement in its 3rd 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 standard 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 4th 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.

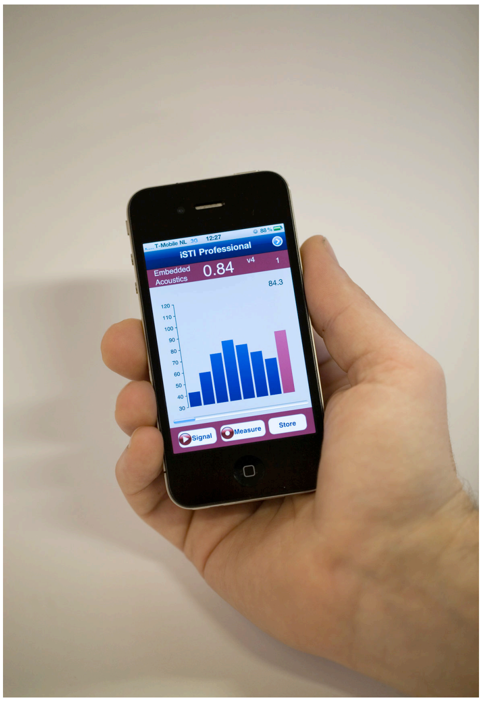

Figure 4. iPhone app for performing 4th edition-compliant STIPA measurements (2011)

2.3 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.

3 CURRENT CHALLENGES

3.1 STI-related ongoing research topics

STI-related research has been ongoing at many institutes and businesses worldwide. TNO in Soesterberg (The Netherlands) kept playing a pivotal role until 2010. TNO has indicated that it will no longer maintain a standalone development program on the STI. Instead, it will collaborate with its spin-out company Embedded Acoustics, and the literally hundreds of authors across the globe that have over the years contributed to the body of literature on the STI.

Some 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. Work on other topics is still ongoing. We will highlight three research subjects that are likely to produce ready-to-use results by the next version of the standard.

3.1.1 Speech-based STI

Measuring the STI using real, recorded, speech was something 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 [9].
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 [10] and Payton [11], 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 and proven promising [12]. 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.

3.1.2 Binaural STI

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 [13] to incorporate binaural listening. Similar work has been done [14] 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.

3.1.3 Measuring the “full” STI with modulated noise carriers

Ideally, the STI is calculated from a densely sampled estimate of the Modulation Transfer Function. In practice, a discrete MTF matrix is always used, with modulation frequencies sampled in 1/3 octave bands from 0.63 Hz to 12.5 Hz.
In practice, a sparsely sampled MTF matrix suffices for most applications. For telecommunication channels the STITEL signal, with just one modulation per octave band, has long been used. STIPA features 2 modulation frequencies per octave band, which was validated to give adequate coverage of the MTF matrix for PA applications.
However, there are some applications where the maximum accuracy and reproducibility of the STI model is achieved only when the full MTF matrix is sampled. For instance, this is the case when nonlinear distortion occurs as well as strong, discrete echoes. Only the TNO reference system currently features a fully IEC-compliant measurement mode for full STI measurements. Other measurement systems that report full STI results are based on inverse calculation through an estimated impulse response, which does not deal with nonlinear effects correctly. 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).

3.2 Future-proof approach towards validation and certification

TNO has always maintained infrastructure to validate new STI measuring equipment: the COMCHA communication channel simulator, as well as an array of previously validated measuring devices (RASTI, STIDAS IID) and measuring software. Validation services based on these assets will no longer be offered. In practice, there is no other institute or company that would be capable and willing to take over this service with the same level of confidence, expertise and independence.
Since the mechanics of the STI method are now firmly set in the latest version of the STI standard, it is mainly validation of implementations that is needed. Our proposal is to create an open-source solution. TNO and Embedded Acoustics will collaborate in creating 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.

3.3 Guidelines for dealing with measurement statistics

Another open issue is the question of dealing with the statistics associated with STI measurements. As with any physical measurement, there is a certain degree of inherent uncertainty (measuring noise) associated with each STI measurement. Repeating the same measurement may result in a value that differs up to 0.02 on the 0-1 STI scale. This is usually not a problem, professionals are normally trained to deal with this type of uncertainty.
In practice, complications associated with the statistics surrounding measuring strategies do arise, involving questions such as:

- How many measuring positions are needed for a certain space?
- How many repetitions per measuring position are needed?
- What criteria should be maintained for discarding measurement points considered to be erroneous, or statistical outliers?
- Given the fact that statistical variation is to be expected in any measurement, how should the "hard" STI boundaries used in standards be dealt with?

One issue that is observed in the practice of PA certification is that hard limits on the STI (such as the general requirement that the STI should always exceed 0.45) is not always fair. Between two competing systems, the one that shows the best average performance may loose out to the other because of a single position at which the STI is somewhat lower. This is sometimes dealt with by accepting lower values at a fixed percentage of measurement points. However, this strategy requires that many measurement points are selected, and according to a predetermined selection regime. Otherwise the results are easily manipulated through the choice of the measurement grid. What is needed here, is to collect good practices from the experience of the STI community, find the common denominator from these practices, and formalize these in a standard.

4 ROADMAP FOR FURTHER DEVELOPMENT OF THE STI

Despite the fact that TNO has lowered its ambitions in sponsoring and coordinating further development of the STI, our expectation is that the rate at which the STI is adjusted to changing demands does not have to decrease. However, the paradigm does have to change – institutes and companies need to find modes to collaborate that are more effective and efficient than seen (in the context of the STI) until now. What we propose comes down to:

- Validation strategies (and infrastructure) based on a freely distributed STIPA database, as described in section 2.2 of this paper;

- Collaboration on future applied research aimed at improving the STI method, based on a commonly accepted roadmap.

It is up to the STI community to develop such a roadmap. A first proposal, with no other intent then to kick off a discussion, has been drafted by the authors of this paper, and is presented in Figure 5. This roadmap includes the topics and issues addressed in this paper.

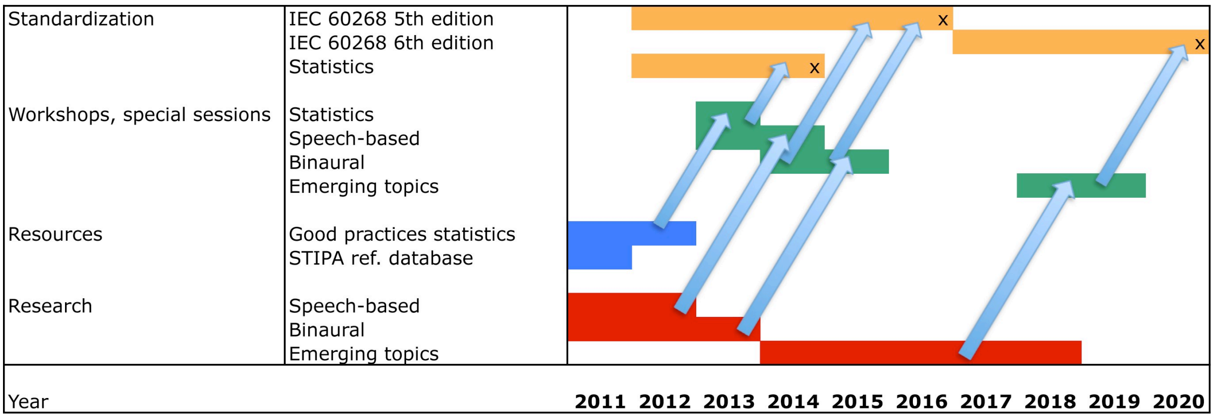

Figure 5. Proposed roadmap (outline) for further development of the STI in the coming decade.

The above roadmap was first presented at the 2011 AES convention in London, and was well received. The first step has meanwhile been taken; the website www.stipa.info has been launched. This is intended to be the portal for the STI community for the coming years. Registered users can share information, articles and files. A first, small database of test signals (intended for verification and validation of STIPA devices) is already downloadable via the website. This database will be expanded soon.

5 SIGNS THAT THE FUTURE IS BRIGHT…

When earlier this year, the 4th edition of IEC-60268-16 was published, hard- and software vendors proved quick to update their products – or at least start the process of updating. This is encouraging; it shows that the market is quick to respond to changes. In fact, there are signs that companies that are newcomers to the STI are planning to release STI-related products soon.

Another sign that the STI is doing well is the interest in specialized STI courses. In response to the initial presentation of the roadmap in Fig. 5, a need was expressed by several parties to organize specialized STI courses for acoustic consultants and engineers. Currently, two courses are foreseen to take place in 2012, as a joint venture between Peter Mapp & Associates and Embedded Acoustics; responses are very positive.

Finally, 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 are two conditions for the STI to develop further: there has to be a community keen on continued use of the STI, and there have to be parties interested in supporting further development and dissemination of the STI. We find that both conditions are met; in fact, the communities of developers and users are overlapping and merging at an increasing rate. 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

6 REFERENCES

1. Houtgast, T., and Steeneken, H.J.M. Evaluation of speech transmission channels by usingartificial signals. Acustica 25, 355-367. (1971).

2. Kryter, K.D. Methods for the calculation and use of the articulation index. J. Acoust. Soc.Am. 34, 1689-1697. (1962)

3. French, N.R., and Steinberg, J.C. Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am. 19, 90-119. (1947)

4. Steeneken, H. J. M., and Houtgast, T. A physical method for measuring speech transmission quality. J. Acoust. Soc. Am. 67, 318–326. (1980)

5. Steeneken, H.J.M., and Houtgast, T. Mutual dependence of the octave-band weights in predicting speech intelligibility. Speech communication, 1999, vol.28, 109-123. (1999)

6. Steeneken, H.J.M., and Houtgast, T. Phoneme-group specific octave-band weights in predicting speech intelligibility. Speech Communication, 2002, vol. 38. (2002)

7. Steeneken, H.J.M., and Houtgast, T. “Basics of the STI measuring method”. In: Past, Present and Future of the Speech Transmission Index, (ed. S.J. van Wijngaarden), TNO Human Factors: Soesterberg, the Netherlands, 125-37. (2002)

8. Jacob, K., Steeneken, H.J.M., Verhave, J., and McManus, S. Development of an accurate, handheld, simple-to-use meter for the prediction of speech intelligibility. Institute of Acoustics, Proc. Reproduced Sound 17, Stratford-upon-Avon. (2001)

9. Steeneken, H.J.M., and Houtgast, T. The temporal envelope spectrum and its significance in room acoustics. Proc. 11th ICA, Vol. 7, 85-88. Paris. (1983)

10. Drullman, R. Temporal envelope and fine structure cues for speech intelligibility. J. Acoust. Soc. Am. 97, 585-592. (1995)

11. Payton, K.L, Braida, L.D., Chen, S., Rosengard, P. and Goldsworthy, R. “Computing the STI using speech as a probe stimulus”. In: Past, Present and Future of the Speech Transmission Index., (ed. S.J. van Wijngaarden), TNO Human Factors: Soesterberg, the Netherlands, 125-37. (2002)

12. Van Gils, B.J.C.M and Van Wijngaarden, S.J. Objective Measurement of the Speech Transmission Quality of Vocoders by Means of the Speech Transmission Index. NATO RTO symposium RTO-MP-HFM-123 “New directions for improving audio effectiveness”. Amersfoort, The Netherlands. (2005)

13. Van Wijngaarden, S.J. and Drullman, R. Binaural intelligibility prediction based on the speech transmission index. J. Acoust. Soc. Am. 123, 4514–4523. (2008)

14. Beutelmann, R., and Brand, T. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 120, 331–342. (2006)