RECENT ADVANCES IN STI MEASURING TECHNIQUES

SJ van Wijngaarden TNO Defence, Security & Safety, Soesterberg, The Netherlands
JA Verhave

1 INTRODUCTION

The Speech Transmission Index (STI) has been the subject of (almost) continuous development and refinement since its introduction in the 1970s. Scientific papers tend to cite Steeneken and Houtgast's 1980 paper[1] as the original reference for the STI method. The version of the STI detailed in that article corresponds with the first edition of the IEC-standard on the STI[2]. However, Steeneken and Houtgast's work on the STI started years before that paper; in fact, they first used the term Speech Transmission Index in their 1971 paper in Acustica[3].
Since 1971, each refinement of the STI was validated by means of listening tests, and published in journals and conference proceedings. Major improvements of the STI were consolidated by incorporating these in revised editions of IEC-standard 60268-16. Three editions of this standard have appeared so far (1988, 1998 and 2003), and a fourth edition is expected to appear within the next few years. To date, the process of continuous improvement of the STI is ongoing.
The current paper aims at providing an overview of how the STI has been evolving, up to the most recent advances that are currently being published (and may be considered candidate-modifications for the next version of the standard). Each new extension of the STI method has essentially served the same purpose: widen the application range of the STI by removing limitations of the model. Models are simplifications of reality; the STI is no exception. The method can only be used in cases where the applied simplifications are allowable. Despite the fact that the limits of the model have always been indicated in publications and standards, users of the STI method have shown a consistent tendency to apply the method *beyond* its limits, simply because their applications are beyond these limits. As a result, they may obtain inaccurate intelligibility predictions. A major purpose of this paper is to help STI users to decide on which version of the STI is required for their applications, in order to obtain the most meaningful and accurate results possible.
Section 2 of this paper describes the common characteristics of all published versions of the STI, including similarities and differences in comparison to other speech intelligibility predictors, such as the Articulation Index and the Speech Intelligibility Index. Section 3 outlines the major differences between different versions of the STI in terms of the applied calculation procedures (algorithms). Section 4 explains the differences between different test signals. Section 5 gives an overview of current and recent scientific research in relation to the STI, that may influence future versions of the STI standard. Section 6 provides an overview of main conclusions.

2 COMMON CHARACTERISTICS OF ALL VERSIONS OF THE STI

2.1 STI principles

The STI is rooted in articulation theory and is strongly related to the articulation index[4] (AI). There are, however, fundamental differences between the AI and the STI. Perhaps the most important difference is that the AI is based on the (frequency-band specific) speech-to-noise ratio, and the STI on the Modulation Transfer Function (MTF). In some specific conditions these approaches are

computationally equivalent, but in others (such as any conditions featuring reverberation or echoes) the MTF-based approach can predict effects on intelligibility that the AI cannot. At the base of any STI measurement is an MTF measurement; a basic understanding of the MTF helps in understanding the possibilities and limitations of the STI.
The MTF quantifies how the intensity envelope of the speech signal is affected by a channel. In speech, the signal intensity varies with time. Very fast intensity variations, which define the temporal *fine structure*, occur due to the rapid succession of glottal pulses and noise-like sound produced along the vocal tract of a speaker. These fast fluctuations, although often altered during transmission across speech transmission channels, are normally not the ones determining the overall intelligibility. Intelligibility depends on the *slower* intensity fluctuations (say, 1 to 15 Hz in frequency). These slower fluctuations result from geometric transformation patterns of the vocal tract (articulatory gestures) which people apply while forming phonemes, syllables and words.
Whether or not these modulations are preserved by the channel determines whether or not the information represented by the speech signal is preserved. An example of an intensity envelope is given in Fig. 1.

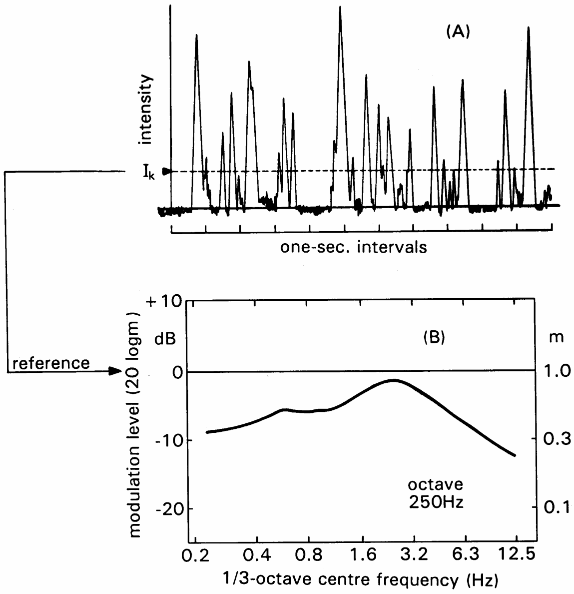

Figure 1. Envelope function (panel A) of a 10s speech signal within the 250 Hz octave band. The corresponding envelope spectrum (panel B) is normalised with respect to the mean signal intensity (I_k).

Figure 1 shows the intensity fluctuations (the envelope) of a speech sample in the time domain (A) as well as in the frequency domain (B). Loss of any of the modulation frequencies in the so-called envelope spectrum shown in Fig. 1B implies loss of information.
The MTF can be thought of as the envelope spectrum at the end of a channel under test, divided by the envelope spectrum at the beginning of the channel[i]. If these envelope spectra are the same, then the MTF is equal to 1, and no information has been lost. If there are no modulations left at the end of the channel, then the MTF is zero, and all speech information is fully lost.

Figure 1 represents the 250 Hz octave band. Because different frequency bands contribute differently to overall intelligibility, the STI is based on MTF measurements in seven octave bands (125-8000 Hz). A full STI measurement involves measuring the MTF for 14 different modulation frequencies (the horizontal axis in Fig. 1B) and 7 octave bands. This 14x7 matrix is converted into a single digit, the STI, by means of a set of algorithms which incorporate perceptual and linguistic knowledge. It is this MTF-to-STI conversion that has mostly been improved over time, together with the level of sophistication of the applied test signals.

[i] Although this statement is conceptually correct for linear time-invariant systems, this is normally not a good way to measure the MTF in practice.

All versions of the STI published between 1980 to 2006 have the following in common:

- Modulation frequencies between 0.63 and 12.5 Hz are considered to contribute to intelligibility (although not necessarily all are measured). All modulation frequencies are considered equally important.
- The analysis is performed on separate octave bands, ranging from 125 Hz to 8 kHz
- These octave bands differ in their relative importance to the overall intelligibility. This is reflected in the STI algorithms by weighting functions. The STI is a weighted sum of MTF values from each of the octave bands

2.2 Relation to Articulation Index (AI) and Speech Intelligibility Index (SII)

Since the STI and the Articulation Index[4] (AI) have common theoretical foundations, there are numerous similarities. However, the two models have evolved differently; the Speech Intelligibility Index (SII)[5], which is the current successor to the AI, targets different populations and applications than the STI. For many applications, acousticians have significant freedom to choose between the two models (each having its own advantages and drawbacks). For some applications, the advantage of one model over the other is clear and significant.

The best way to address the differences between the two models is by direct comparison between the current versions: the SII according to the 1997 ANSI standard[5] and the STI according to IEC 60268-16 3rd edition[6].

First off, it should be noted that not all differences are immediately apparent when comparing the text of both standards. Both appear very similar indeed: both produce an intelligibility metric on a 0 to 1 scale, based on weighted contributions from a range of frequency bands. It requires closer study to spot the differences.

The first major difference is crucial for the type of application that may be addressed, but is easily missed at first sight: the SII has to be *calculated*[ii], while the STI can either be calculated or *measured directly*. The procedures for obtaining the SII require that measurement data are available as input (speech spectra, noise spectra, overall signal levels). The same data could also be used to calculate the STI. Moreover, the STI can also be determined by an *in situ* 10-15 second measurement, producing a near-instant intelligibility prediction.

The SII and STI also differ in the types of signal distortions they are sensitive to. As mentioned in section 2.1, the STI is sensitive to time-domain distortions such as reverberation and echoes, whereas the SII is not. The same is true for various nonlinear distortion types. On the other hand, the SII incorporates the decrease of intelligibility that occurs with high vocal effort, while the STI does not. Also, the SII offers greater flexibility in the type of frequency filtering (allowing better frequency resolution) and different options for the frequency weighting functions, depending on the amount of redundancy provided by the linguistic content of the speech signal.

Many differences in design follow from a difference in general philosophy: the SII aims at predicting speech intelligibility (including specific and individual effects of speakers and listeners), whereas the STI quantifies speech transmission quality (the degree to which the channel affects intelligibility). The difference may appear subtle, but it has fundamental consequences. For instance, the SII assumes different values for different kinds of speech utterances (in terms of linguistic redundancy); the STI depends only on the channel. In other words, a single channel always produces the same STI value, while the SII may vary. This makes the STI more suitable for comparing, checking and benchmarking rooms or channels, while the SII is more suitable for finding out the true level of intelligibility experienced by an individual in a well-defined situation.

As a consequence of all these differences, the SII and the STI have different user populations. For instance, experimental audiologists generally prefer the SII, for its higher frequency resolution and its sensitivity to the intelligibility decrease at high vocal efforts. The SII is also often preferred by those who are interested in comparing effects of different speech materials rather than different channels. The STI is more often preferred by acoustic consultants, who require sensitivity to reverberation, and by telecommunication engineers, who depend on the model being sensitive to nonlinear distortions.

[ii] Section 5.2.3. of ANSI S3.5-1997 shows how the MTF could be estimated using test signals, in a simplified procedure derived from the STI approach. This opens up possibilities to directly measure the SII in an STI-like manner, but the approach is impractical, lacks uniformity, is insensitive to non-linear distortion types, and appears not to be generally used.

2.3 Relation between the STI and other intelligibility predictors

Intelligibility prediction metrics can be broadly divided into two categories: relatively complex predictors including explicit and sophisticated perceptual and cognitive modelling, and simpler metrics that are easier to measure and understand and are therefore accessible to greater populations of acousticians. The STI and SII both fall into the first category, although the STI leans towards the ease-of-use which is the benefit of the second category, while the SII more dominantly possesses the flexibility and scientific rigor that is the benefit of the first category.
Another example of the first category (complex perceptual models) is the Speech Recognition Sensitivity model[7], which elegantly works around shortcomings of other models, but has not seen much "field experience" or independent evaluation. Complex models have also been developed to predict speech quality and intelligibility specifically for telecommunication channels (for example, the PESQ model[8]). The added value of the STI, in relation to these models, is the wider applicability (room acoustics *and* telecommunications), combined with its widespread use and third-party evaluations. The fact that various vendors have implemented the STI method in their measuring devices helps in this respect.
The category of simpler metrics includes the Speech Interference Level (SIL)[9], a metric that predicts intelligibility of speech in noise by averaging the speech-to-noise ratio in three octave bands. It also includes various measures based on early-to-late energy ratios derived from impulse responses, such as clarity and definition[10]. These are specifically of interest when investigating reverberant environments. Under the conditions (and for the type of applications) that these measures are intended for, their level of accuracy may approach that of the STI. In more complex situations, the accuracy of the STI far outstrips all simpler metrics.

3 DIFFERENCES BETWEEN ALGORITHMS OF SUCCESSIVE STI VERSIONS

3.1 First standardised version (IEC 268-16, 1st edition)

This first edition of the IEC-standard is based on one (simplified) type of test signal called RASTI (explained in section 4). The computation algorithms are based on Steeneken and Houtgast's 1980 paper[1] and apply equally to other test signals as well as RASTI. All frequency bands are considered statistically independent; this results in the fact that channels with irregular frequency transfer functions (frequency "gaps"), or extremely narrow-band channels, will produce inaccurate intelligibility estimates.
Auditory masking slopes are uniformly assumed to be -35 dB/oct. In reality, these slopes flatten out at high listening levels, leading to increased masking effects. As a result, this version of the STI will be an inaccurate intelligibility predictor at high speech and noise levels - say, in excess of 90 dB(A).

3.2 Mutual octave band dependency (IEC 60268-16, 2nd edition)

The second edition of the IEC-standard introduced several new types of test signal and provided more complete descriptions of the various algorithms. New algorithms are introduced that take the mutual dependency between octave bands into account, fixing the inaccuracies for channels with irregular frequency transfer functions. Different versions of the algorithm for male and female speech are introduced, to take the lack of energy in the lower frequency regions of the female voice into account. An absolute reception threshold is introduced to reflect the lack of human speech reception capabilities at low sound levels.

3.3 Level-dependent masking (IEC 60268-16, 3rd edition)

By introducing level dependent masking, accuracy of intelligibility predictions at higher sound levels is enhanced. Also, a new test signal is introduced (STIPA, see also section 4), and limitations of the STI method, as well as the relation to subjective intelligibility measures, are addressed more explicitly.

3.4 Known current limitations

Some conditions may produce inaccurate results with any of the current versions of the STI; they are simply beyond the scope of the STI. The major limitations are:

- Strongly fluctuating noise produces inaccurate results. At the lower end of the STI scale intelligibility tends to be overestimated; higher intelligibility levels are underestimated.
- Centre clipping (nowadays rare, but used to occur with carbon microphones and poorly adjusted low-class amplifiers) leads to an overestimation of intelligibility.
- If sources of noise and speech are clearly spatially separated, intelligibility is underestimated since effects of binaural hearing are not taken into account
- Amplitude compression may lead to inaccurate intelligibility predictions
- Digital vocoders (voice compression techniques used for instance in satellite phones and voice-over-IP) may produce overestimated intelligibility results.

4 STI TEST SIGNALS

Whether or not accurate results are obtained does not only depend on the version of the calculation algorithms that is used; it also depends on the suitability of the test signal. An overview of available test signals is given in table I.

Table I. Overview of STI test signals, with the distortion types these are designed for, required measuring time and the version of IEC 60268-16 in which the signal type was introduced. For each signal type, the number of octave bands and the number of modulation frequencies per band are also indicated (e.g., 7 x 2 means 7 octave bands, 2 modulation frequencies per band).

Distortion type **Signal Type** (octaves x modulation frequencies)	*Band-pass limiting*	*Non-Linear Distortion*	*Reverberation Echoes*	*Measuring time*	*Introduction in IEC 60268-16*
STI-14 ('full STI') (7 x 14)	yes	yes	yes	15 min	2nd
STI-3 (7 x 3)	yes	yes	condition dependent	4 min	2nd
STITEL (7 x 1)	yes	condition dependent	condition dependent	15 s	2nd
STITEL14 (7 x 14, by 14 consecutive STITEL signals)	yes	condition dependent	condition dependent	14 x 15 s (3.5 minutes)	-
RASTI (2 x 4/5)	no	no	yes	15 s	1st
STIPA (7x2)	yes	condition dependent	yes	30 s	3rd

The term 'condition dependent' is used to indicate that the corresponding test signal type may or may not produce accurate results, depending on the exact distortion type. For example, STITEL *can* be used in reverberant environments, provided that the reverberation time is not too dependent on frequency. Similarly, STIPA can be used for PA systems that produce non-linear distortion components, unless the signal is severely clipped in various frequency bands.

5 RECENT DEVELOPMENTS

5.1 Binaural STI

Classically, effects of binaural hearing are not included in STI measurements, if only because these measurements are based on a single microphone instead of two ears. In highly reverberant environments, or especially if speech and noise are coming towards the receiver from different directions, the standard STI may produce an underestimation of intelligibility. A method has recently been proposed[11] to calculate a binaural STI, by combining a cross-correllogram model of binaural hearing with the STI. Instead of a single microphone, an artificial head needs to be used, followed by a two-channel STI analysis.

Binaural hearing offers intelligibility benefits over monaural hearing for at least two reasons. Firstly, because of acoustic shielding effects by the head and torso, effective speech-to-noise ratios normally differ between the ears. This offers the possibility to simply focus on the ear with the best signal quality. The second effect is based on our ability to process interaural time difference; a neural cross-correlation process enhances our “internal” speech-to-noise ratio. These two effects are incorporated into the binaural STI by simple best-ear selection rules for all octave bands, combined with a cross-correlation approach in the 500, 1000 and 2000 Hz octave bands (covering the frequency region in which humans are able to process interaural time differences). This binaural STI was validated through Consonant-Vowel-Consonant (CVC) intelligibility tests, as is customary after modifications of the core STI algorithms. Results from these validation tests are given in Fig. 2.

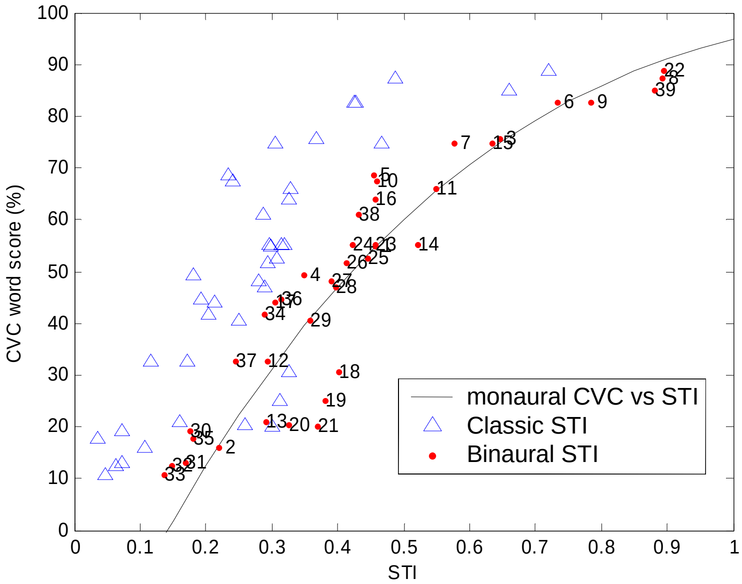

Figure 2. CVC word score (7 subjects) as a function of the binaural STI as well as the monaural STI. Test conditions were selected to be “difficult” for the standard STI; the conditions include anechoic conditions (1-14), a cathedral environment (15-21), a classroom (22-32) and a listening room (33-39). The standard “monaural CVC vs STI” reference curve is also given.

As Fig. 2 shows, the standard STI gives an underestimation of intelligibility in these (selected) binaural listening conditions. The binaural STI algorithm comes much closer to the reference curve. In conclusion, if accurate intelligibility predictions are needed in conditions that are susceptible to significant binaural processing advantages, it will be worthwhile to use the proposed binaural STI algorithm.

5.2 Using speech as a test signal

The fact that the STI uses artificial test signals is the main reason why channels that feature vocoders lead to inaccurate results; digital vocoders use compression techniques based on knowledge of the inherent structure of speech signals. When artificial signals (such as the modulated noise signals used for the STI) are processed through vocoders, the audio becomes distorted: the vocoder tries to reproduce the sound as if it were speech, while it is really noise instead (which these systems normally try to suppress). The obvious solution is to measure the STI using real speech as a test signal. Unfortunately, this is more complicated than it may seem. One of the reasons that STI measurements can be done quickly and accurately is that known, stable modulation frequencies are used in the test signals. The envelope spectrum of natural speech is not as stable, and includes simultaneous contributions from all modulation frequencies in the relevant

frequency range. Moreover, noise also contains spurious, speech-like modulations. A speech-based analysis algorithm needs to separate modulations due to noise from modulations due to speech.
The idea to use speech as a test signal is certainly not new; Steeneken and Houtgast explored it before any of the current signals were even designed[12]. However, only due to advances in digital signal processing capabilities has such an approach become feasible. Various methods have been proposed[13,14,15]. A recent refinement of these methods[16,17] has been shown to produce reliable results with vocoders, as well as with a range of more conventional channels.
For conventional channels (featuring noise, reverberation, peak clipping and bandwidth limiting), the measurement error margin of the STI can be equal to the error margin of the artificial signals such as STIPA (that is, approx. 0.02) provided that the test signal length is at least 40 seconds. For vocoders, an additional complication arises: vocoders degrade the fine structure of speech to such a degree that intelligibility is affected. This is uncommon among speech communication channels; the STI is not sensitive to changes in the fine structure, and misses their effect. To investigate the consequences, speech processed through 9 narrow-band digital vocoders (collected during a NATO vocoder contest for digital military radio applications) was subjected to speech-based STI measurements, as well as subjective CVC intelligibility tests. Results are shown in Fig. 3.

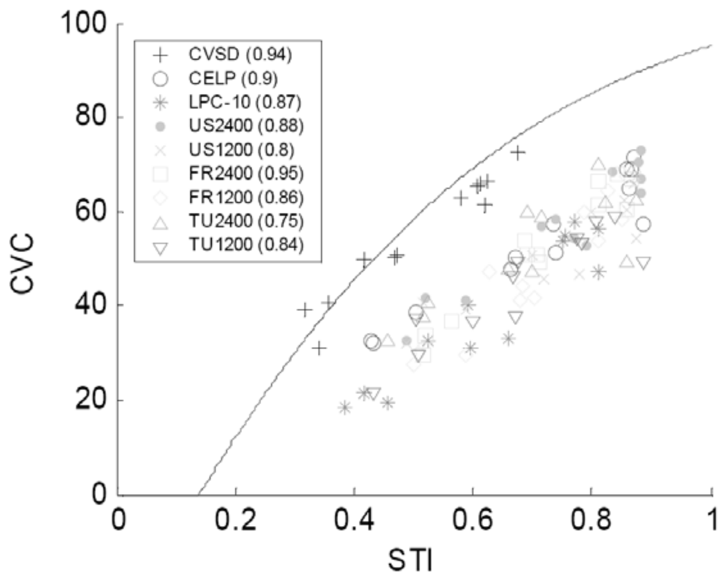

Figure 3. CVC word score (4 subjects) as a function of the speech-based STI for 9 vocoder types. The standard “monaural CVC vs STI” reference curve is also given.

With the exception of CVSD, which is really a waveform coder instead of a vocoder and does not affect the fine structure of speech, the data points deviate systematically from the reference curve. However, a monotonically increasing regression line can be drawn through the data points: the STI *does* increase with increasing intelligibility as desired, just not according to the normal function.
This can be accounted for in two ways: the interpretation of the STI can be adjusted, or a numerical compensation can be applied. By subtracting 0.30 from any STI value obtained with a vocoder, the data points are shifted left in Fig.3. The spread around the reference curve is then acceptable, and of similar magnitude as for other conditions.
In conclusion, it *is* entirely possible to carry out STI measurement with vocoders using speech as a test signal. This requires that results are adjusted post-hoc to take fine structure effects into account. Since the same numerical adjustment (0.3) suffices for a range of different vocoder types, this can be done without limiting the applicability or reliability of the method.

5.3 Special populations: non-native listeners

For normal populations of talkers and listeners, it is known how to translate an STI value into a qualification of intelligibility: the relation between STI and subjective intelligibility test scores (such as CVC, DRT, MRT and SRT) is known. For most purposes, it suffices to derive a global intelligibility qualification from the STI by using a standard table, reproduced here as table II.

Table II. Translation of the STI into an intelligibility qualitication label.

Qualification	*Bad*	*Poor*	*Fair*	*Good*	*Excellent*
STI range	0 – 0.30	0.30 – 0.45	0.45 – 0.60	0.60 – 0.75	0.75 – 1

However, if the population is not normal, the relation between STI and intelligibility may be different, and a different table of qualifications is needed. For instance, a listener suffering from a hearing impairment will experience a reduction in intelligibility compared to an otologically normal reference population. He may experience “poor” intelligibility under conditions normally rated “good.” Something similar is true for populations of non-native listeners, for whom the language spoken is not their mother tongue. Because they have learned the language at a later age, their command of the language is lower than for natives, resulting in a reduced level of speech intelligibility. The magnitude of this effect depends on their level of proficiency[18]. Given any population of non-native subjects, the reduction in intelligibility can be measured through listening tests.

To predict true intelligibility for a known population of non-natives from STI measurements, specific qualification tables need to be constructed. This can be done following the procedure[19] shown in Fig. 4.

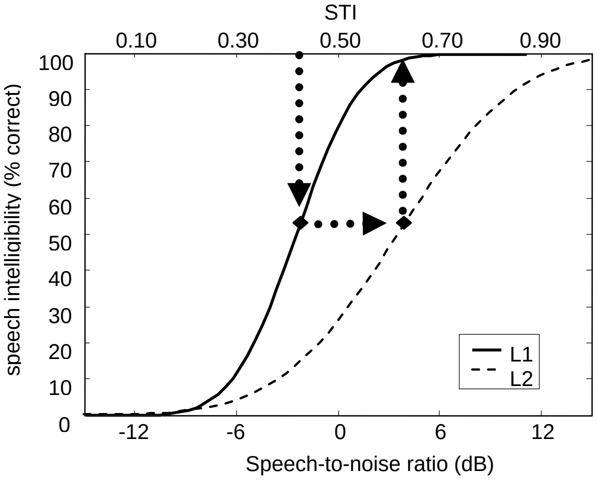

Figure 4. Procedure for deriving new STI qualification category boundaries from native (L1) and non-native (L2) psychometric curves. These curves need to be determined by carrying out speech-in-noise intelligibility experiments with a native reference population as well as the non-native population under consideration. Any “native” STI value can then be translated into a non-native equivalent, corresponding to the same intelligibility, by following the arrows.

Table III contains adjusted intelligibility qualification tables for three groups of non-natives, obtained from experimental data[19] through the process of Fig. 4. For low-proficiency non-native listeners, good or excellent intelligibility cannot be achieved.

Table III. Adjusted intelligibility qualification tables for three categories of non-natives

STI label category boundary	Standard STI	Nonnative cat. I (experienced, daily L2 user)	Nonnative cat. II (intermediate experience and level of L2 use)	Nonnative cat. III (new learner, infrequent L2 user)
bad –poor	0.30	0.33	0.38	0.44
poor – fair	0.45	0.50	0.60	0.74
fair – good	0.60	0.68	0.86	impossible
good - excellent	0.75	0.86	impossible	impossible

6 CONCLUSIONS

We realise that this paper may leave the reader somewhat confused and uncertain about the reliability of the STI. If the STI is a trustworthy, standardised measure, then why does it need constant improvement? And why should it matter which version we use? Fortunately, reality is reassuring: in most cases, anybody who picks out a (certified) STI measuring device and starts using it, maybe even without reading the manual, is very likely to immediately get meaningful and accurate results. If the device was bought recently, chances of good results are even better. Even if the STI is applied inappropriately, and accuracy of results is not guaranteed, the resulting error may still be acceptably small. However, to get optimally reliable results *all* the time, and to be true to the professional standards acousticians strive to uphold, one should be aware of the limitations of the method. For this reason, and to stress the need for continuous development, this paper was written.

Continuous development and improvement has already resulted in the following: the newer the version of the standard your device complies with, the wider the range of applications it can be used for without restrictions.

To conclude, table IV provides an overview of STI versions and test signals that are preferred, acceptable and unacceptable for various applications and conditions.

Table IV. Suitability of test signals and different STI versions for various conditions and applications, ranging from – to ++. The symbol – means that corresponding should not be used; + indicates and acceptable choice, ++ is the preferred choice.

Application	*Full STI (STI-14)*	*STITEL*	*STIPA*	*RASTI*	*1st edition*	*2nd edition*	*3rd edition*
Assessing suitability of room acoustics for speech communication (no electronic amplification)	+	+ / - Depends on RT	+	+	+	+	++
Evaluating PA and voice warning systems	+	+ / - Depends on reverb & distortion	++	-	-	+	++
Evaluating telecommunication channels (phone, radio)	+	++	+	-	-	+	++
Speech levels at listener positions higher than 90 dB (A)	+	+	++	-	-	-	++
Speech levels at listener positions lower than 40 dB (A)	+	+	+	-	-	+	++
Difference between male and female voices needs specific attention	+	+	-	-	-	+	++
Channel features amplitude compression	+	+	+	-	-	+	++
Significant parts of the listener population are non-native	Standard STI qualification labels (categories) are inappropriate. STI can be measured as usual, but should be interpreted differently				+	+	++
Centre clipping	No method currently available				-	-	-
Strongly fluctuating noise	Work with average of several STI measurements				-	-	+
Extreme reverberation times (>5 s), or speech and noise clearly spatially separated	Currently standardised methods are inaccurate Use of binaural STI is recommended				-	-	-
Channels that do not permit artificial test signals, such as vocoders	Currently standardised methods are inaccurate Use of a speech-based STI test signal is recommended				-	-	-

7 REFERENCES

1. H.J.M. Steeneken and T. Houtgast. 'A physical method for measuring speech transmission quality'. J.Acoust.Soc.Am. 67, 318-326. (1980).
2. International Electrotechnical Commission IEC 268-16. 'The objective rating of speech intelligibility in auditoria by the 'RASTI' method'. Geneva, Switzerland (1988).
3. T. Houtgast and H.J.M. Steeneken, 'Evaluation of speech transmission channels by using artificial signals'. Acustica 25, 355-367. (1971).
4. ANSI S3.5-1969, American national standard; methods for the calculation of the articulation index, American National Standards Institute, New York. (1969).
5. ANSI S3.5-1997, American national standard methods for the calculation of the speech intelligibility index, American National Standards Institute, New York. (1997).
6. International Electrotechnical Commission IEC 60268-16 3rd edition. 'Sound system equipment – Part 16: Objective rating of speech intelligibility by Speech Transmission Index'. Geneva, Switzerland. (2003).
7. H. Müsch and S. Buus. 'Using statistical decision theory to predict speech intelligibility. I. Model structure'. J.Acoust.Soc.Am. 109, 2896–2909. (2001).
8. J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. 'PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model'. J. Audio Eng. Soc., vol. 50, pp. 765-778. (2002 Oct.).
9. International Organization for Standardization ISO 9921. '"Ergonomics - assessment of speech communication". (2003).
10. H. Kuttruff. 'Room Acoustics' Spon Press, UK. (1991).
11. S.J. van Wijngaarden and R. Drullman. 'Development of a binaural speech transmission index (S).' J.Acoust.Soc.Am. 119(5). 3442. (2006).
12. H.J.M. Steeneken and T. Houtgast., The temporal envelope spectrum and its significance in room acoustics, Proc. 11th ICA, Vol. 7, 85-88. Paris (1983).
13. C. Ludvigsen, C. Elberling, G. Keidser and T. Poulsen. 'Prediction of intelligibility of nonlinearly processed speech'. Acta Otolaryngol Suppl., 469, 190-5. (1990).
14. R. Drullman, J.M. Festen and R. Plomp. 'Effect of reducing slow temporal modulations on speech reception'. J. Acoust. Soc. Am., 95, 2670-2680. (1994).
15. K.L. Payton, L.D. Braida, S. Chen, P. Rosengard and R. Goldsworthy (2002). Computing the STI using speech as a probe stimulus. Past, Present and Future of the Speech Transmission Index., (ed. S.J. van Wijngaarden), TNO Human Factors: Soesterberg, the Netherlands, 125-37. (2002).
16. 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. NATO-RTO symposium RTO-MP-HFM-123 "New directions for improving audio effectiveness". Amersfoort, The Netherlands. (2005).
17. R. Drullman and S.J van Wijngaarden. 'New directions for a speech-based speech transmission index (S).' J.Acoust.Soc.Am. 119(5). 3442. (2006).
18. S.J. van Wijngaarden, H.J.M Steeneken and T. Houtgast 'Quantifying the intelligibility of speech in noise for nonnative listeners.' J.Acoust.Soc.Am. 111(4), 1906–1916. (2002)
19. S.J. van Wijngaarden, A.W. Bronkhorst, T. Houtgast and H.J.M Steeneken. 'Using the Speech Transmission Index for predicting non-native speech intelligibility' J.Acoust.Soc.Am. 115(3), 1281–1291. (2004).