

Basics of the STI-measuring method

Herman J.M. Steeneken and Tammo Houtgast

Preface

In the late sixties we were asked to perform range measurements for VHF-radio systems. These measurements should make use of (subjective) intelligibility tests. The effort required for this project was enormous. This was due to the number of individual parameters included in the test, but also by the time consuming subjective intelligibility measurements. Therefore, we initiated the use of objective testing in order to *predict* the intelligibility by simple physical measurements. This first step was very much appreciated and resulted into an objective intelligibility measure: the Speech Transmission Index (STI). The measurement of the STI was performed with a simple analogue real-time measuring system (STIDAS-I).

Further developments have led to a robust method that produced an accurate prediction of the intelligibility for many types of transmission channels and in room acoustics: the STI method. This procedure was also realised in a specific measuring device (STIDAS-II, 1978). Twenty-five of these devices, which were based on specific hardware and a PDP 11-03 computer, were in use all over the world.

As a spin-off, a screening device for measurement of the STI in auditoria was developed in 1979. The RASTI method (Room Acoustical Speech Transmission Index) is defined in an former IEC recommendation IEC 268-16. Several companies built specific hardware for the measurement of RASTI, or incorporated STI-related measures in their own systems.

The accuracy of the STI method has been improved ever since and has been extended to predict the intelligibility for both male and female speech. The application is not restricted to specific hardware but has been implemented in a software package. The use of the STI-method has grown steadily over the past years. Many standards and recommendations on transmission quality include the STI procedure (ISO9921, IEC 60268-16). In relation to this the RASTI system is often used for assessment of communication systems including deteriorated sound sources for which the RASTI method is *not* designed. For this purpose the STI-PA is recently designed. This system is applicable for public address systems and accounts correctly for the distortions that are related to public address. The test signals are provided on a CD and a specific hand-held analyser performs the analysis.

This overview describes the principles underlying the STI method and gives a detailed description of the use of the method, the diagnostics, and examples of a number of applications.

1 Introduction

Speech is considered to be the major means of communication between people. In many situations the speech signal we are listening to is degraded, and only a limited transfer of information is obtained. This may be due to factors related to the speaker, the listener, and the type of speech, but in most situations it is due to limitations imposed by the transmission of the speech signal from the speaker's mouth to the listener's ear. The purpose of the measuring method described in this overview is to quantify these limitations and to identify the physical aspects of a communication channel that are primarily related to the intelligibility of the speech signal passed through such a channel. During transmission, degradation may occur that results in a decrease of the information content¹ of the speech signal such as: limitations of the frequency range, the dynamic range, and distortion components.

All these aspects have been studied in the literature during the past seven decades. This has resulted in design criteria for transmission channels and in the development of speech quality measures, speech intelligibility tests, articulation tests, and a few diagnostic and objective assessment methods. Three methods of assessment can generally be distinguished:

- (a) subjective measures making use of speakers and listeners,
- (b) predictive measures based on physical parameters,
- (c) objective measures obtained by measurements with specific test signals.

- (a) Subjective tests make use of various types of speech material. All these tests have their specific advantages and limitations mostly related to the speech items tested. Frequently used speech elements for testing are phonemes, words (digits, alphabet, short words), sentences, and a free conversation in combination with quality rating.

- (b) Predictive measures based on physical and perceptual parameters that quantify the effect on the speech signal and the related loss of intelligibility due to for instance: a limited frequency transfer, masking noise, reverberation, echoes, and a non-linear transfer resulting from peak clipping, quantisation, or interruptions.

From the perceptual (listener) point of view, hearing properties, such as frequency resolution, auditory masking, and reception thresholds, also define the intelligibility for a given condition.

One of the first descriptions of a model to predict the effect of a transmission path on the intelligibility of speech was presented by French and Steinberg (1947) and later evaluated by Beranek (1947). This work formed the basis for the so-called Articulation Index (AI), which was described, evaluated and made accessible by Kryter (1962a).

1) Information content: properties of a speech signal that contribute to identification of a speech item (phoneme, word, or sentence).

(c) The objective measurement of speech intelligibility has been studied for many years. Specific measuring devices were developed, improvements were made, and the range of applications extended. Therefore, in the next chapter, an overview is given of these developments during the past forty years.

One such objective method to *predict* the speech transmission quality of an existing communication channel was developed by Houtgast and Steeneken (1971), and Steeneken and Houtgast (1980). This method is based on the application of a specific test signal. The transmission quality is derived from an analysis of the received test signal, and is expressed by an index, the Speech Transmission Index (STI). The STI is based on weighted contribution from a number of frequency bands. For this purpose, the STI uses a fixed bandwidth (octave bands) with a contribution (weighting factor α_k) as indicated in Fig. 1.1.

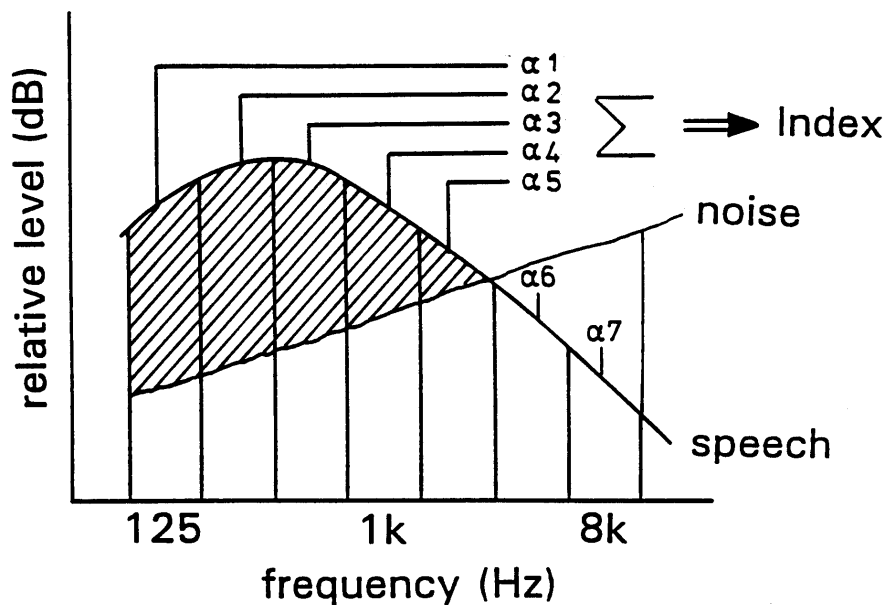


Fig. 1.1 Illustration of the long-term spectrum of a speech signal masked by noise, and the weighted summation to an objective intelligibility prediction.

The STI value is obtained from measurements on the transmission channel in operation, or based on a calculation scheme making use of physical properties of the transmission channel. The STI measurement requires a special test signal from which the *effective* signal-to-noise ratio in each octave band at the receiving side is determined and used for the calculation of the STI. The specific features of this approach are that the test signal design allows an adequate interpretation of many degradations than just a limited frequency transfer and masking noise, for example non-linear distortion and distortion in the time domain. Hence, almost all types of distortion and their combinations that may occur in an analogue or digital (wave-form-based) transmission path are accounted for. However, distortions such as frequency shifts and voiced/unvoiced decision errors that may occur

with certain types of vocoders, are not included in this concept. Up to 1993 the measurement of the STI for telecommunication channel evaluation made use of a specific measuring device (Steeneken and Agterhuis, 1982). Over the years, twenty-five of these devices have been built and have been distributed to many laboratories all over the world. Fifteen years of experience with the development and the application of the STI have shown the need for further improvements, for instance when applied to conditions with a very limited frequency transfer or non-contiguous frequency transfer. Also, effects of speaker variation, the gender of the speaker, and the individual relation with consonant and vowel recognition required further attention.

We were able to improve the STI-model and extend the model with respect to male/female speech, the type of speech being assessed, and speaker variations. Steeneken (1992) describes the results of this study.

2 Overview of objective measuring methods for predicting speech intelligibility

The first description of the use of a “computational method for the prediction of the intelligibility of speech and its implementation in an objective measuring device” was given by Licklider et al. (1959). They described a system that could measure the spectral correspondence between speech signals at the input and at the output of the transmission channel under test, the so-called Pattern Correspondence Index (PCI). This PCI shows a remarkable similarity with the AI (Articulation Index), although the approach is quite different. A spectral-weighted contribution of the similarity between temporal envelopes of the speech signals at the input and at the output of a transmission channel is used for the computation of the PCI. A total of 15 minutes of speech was required for this analysis. The paper reports that the results of a comparison between the PCI and human listener evaluation show a monotonic relation for conditions with an increasing effect of one type of distortion. Contributions of different types of distortion show a "sufficient agreement". Schwarzlander (1959) described the electronic design of the system. Licklider proposed an improvement of the PCI by making use of synthetic signals, physically related to average speech, and with a duration of about one-second for the total measurement of the PCI.

Five years later Kryter and Ball (1964) described a system called the Speech Communication Index Meter (SCIM), which was based on the AI as described by Kryter (1963). The measurements were mainly concentrated on deriving the signal-to-noise ratio within a frequency range of 100-7000 Hz and a dynamic range of 30 dB. The auditory masking corrections according to the AI concept were also included. An evaluation of the system was performed for several types of transmission conditions, including low-pass filtering, noise, frequency shifts, and clipping.

In 1970 we developed a system based on the use of an artificial test signal which was transmitted over the channel-to-be-tested and which was analysed at the output. The test signal was an amplitude-modulated noise signal with a square-wave envelope. Hence the signal level alternated between two values. The difference between these two levels was 20 dB and the switching rate was 3 Hz (Houtgast and Steeneken, 1971). The noise carrier had a frequency spectrum corresponding to the long-term speech spectrum. This was the first approach in which speech-related phenomena, concerning spectral variations and temporal variations, were included in an artificial test signal. The essential point of this approach was that the resulting level variation at the output of a communication system reflects the signal-to-noise ratio, providing a basis for subsequent calculations according to the AI concept. The method was based on measurements in five octave bands (centre frequencies 250 Hz - 4 kHz). The effect of band-pass limiting, noise, peak clipping, and reverberation on intelligibility was included in the test signal concept and in the evaluation procedure. This resulted in an index ranging from 0 - 1, the so-called Speech Transmission Index (STI). A measuring device was developed, based (at that time) on analogue circuits, which could determine the STI within 10 s. It should be noted that this method is different from the STI approach published later and described in chapter 3.

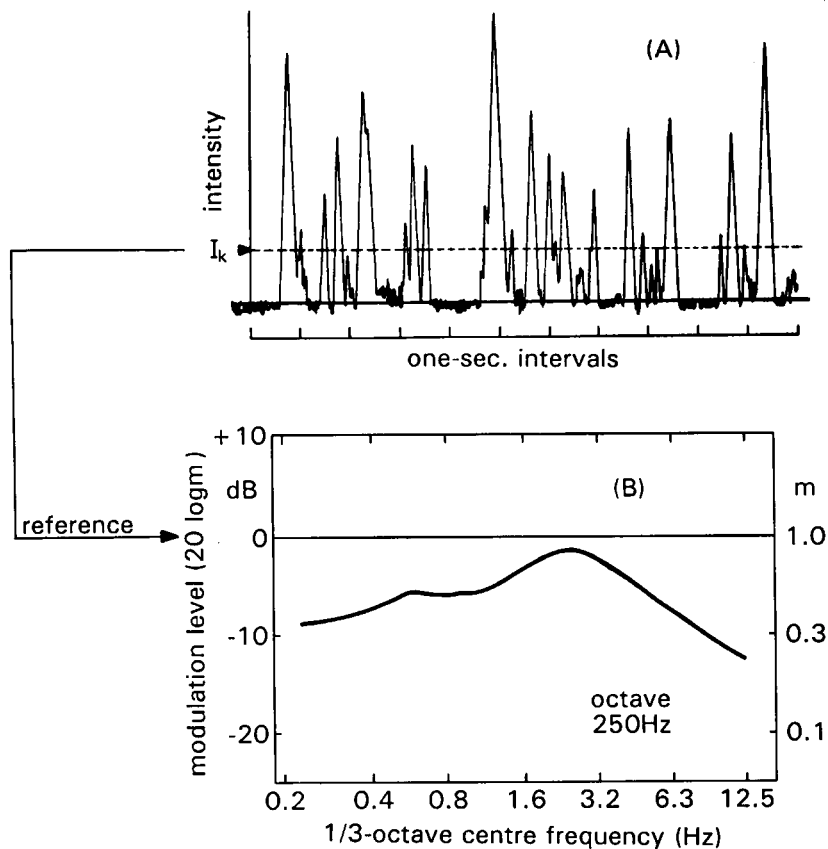


Fig. 2.1 Envelope function (panel A) of a 10s speech signal filtered for the octave band-with centre frequency 250 Hz. The corresponding envelope spectrum (panel B) is normalised with respect to the mean signal intensity (I_k).

The next step was to use a test signal with various modulation frequencies instead of the fixed (3 Hz) square-wave modulation signal. This modulated test signal was based on the measurement of the fluctuations of the envelope of connected discourse (Houtgast and Steeneken, 1971). The envelope fluctuations were determined for separate frequency bands (octave bands). While the envelope function is unique for a certain combination of successive speech sounds, the frequency spectrum of the envelope fluctuations, called the "envelope spectrum" proved to be a stable and reproducible characteristic of running speech (for speech tokens of at least 10 s, see Fig. 2.1). This envelope spectrum (with a frequency range from about 0.2 Hz to 12.5 Hz) was measured in 1/3-octave bands and normalised with respect to the mean level (intensity).

The transfer of these fluctuations of speech by a communication channel can be obtained by comparing the envelope spectra of the same speech signal at the input and at the output of the channel under test (Steeneken and Houtgast, 1973). For that purpose a 60-second segment of natural speech can be used as the test signal. The effect of noise on the envelope

spectrum of speech is independent of the fluctuation frequency, however, this is not the case for distortions in the time domain. Reverberation will act as a low-pass filter for fluctuations and can be predicted for an exponential decay. Since there is a simple relation between the relative decrease of the fluctuations and the actual signal-to-noise ratio, this relation can be used to measure the *effective* signal-to-noise ratio as a function of fluctuation frequency.

Next to the use of natural speech as a test signal, Houtgast and Steeneken (1972) also proposed the use of an artificial test signal, where each relevant fluctuation frequency was tested separately. This resulted in the so-called Modulation Transfer Function (MTF). The (octave-band specific) MTF represents the transfer of the (octave-band specific) *envelope* of a signal between the input and output of a transmission channel.

The method was extensively evaluated for conditions with noise, reverberation, and echoes. The analysis and the generation of the echo conditions, at that time, were performed with a digital (PDP-7) computer, a system with a 1.75 μ s cycle time and 8K-words of memory!

Payne and McManamon (1973) introduced the Speech Quality Measure (SQM) for communication channels. This system was based on the AI concept. The authors mentioned limitations for digital encoding, fading, and non-linear distortion. They remarked "when using the system it should be checked to have none of these distortions present". The test signal was based on 20 tones with frequencies at the mid-point of the 20 frequency bands with "equal contribution to intelligibility" as used for the original AI concept. The paper also proposes the use of mini-computers to perform the analysis and to display the results. No validation was reported.

Steeneken and Houtgast (1980) extended the MTF approach (that had already been validated for channels with noise, echoes, and reverberation) to channels with distortions more specific for communication channels, namely band-pass limiting, noise, non-linear distortion, quantisation errors from digital coders, and reverberation.

Schroeder (1981) developed a mathematical background of the MTF referred to as CMTF. This function is more generic as it also includes the phase transfer. However, this parameter is not used for the STI.

Based on the STI concept, the RASTI method (Room Acoustical Speech Transmission Index) was developed in 1979 (Steeneken and Houtgast, 1979; Houtgast and Steeneken, 1984). This simplified method was especially developed as a *screening* device for applications in room acoustics and restricted to person-to-person communications. The method was standardised in 1988 by IEC 268-16. Notice that the effect of PA-systems on the frequency transfer and possible non-linear distortion was not accounted for.

Quackenbush et al. (1988) gave an overview of "Objective measures of speech quality" especially applied to digital coders. They also evaluated some objective measures, which were mainly based on signal-to-noise ratios.

A major improvement of the STI method, in use since 1980, was achieved in 1992. The additive model on which the AI and STI were based was extended with a so-called redundancy correction. This correction accounts for the correlation of the information content within two adjacent frequency bands of a speech signal. This essential for systems with a very limited frequency transfer (PA systems) and a discontinuous frequency transfer. Also, various extensions were added to the STI measuring procedure such as a separate assessment of male and female speech, the type of speech material used for the prediction of the intelligibility, and a model for the prediction of speaker variations. The results of this study are described by Steeneken (1992) and by Steeneken and Houtgast (1999, 2002a, 2002b).

3 Measurement and calculation of the STI

3.1 Description of the algorithm

The STI is an objective measure, based on the contribution of a number of frequency bands within the frequency range of speech signals, the contribution being determined by the effective signal-to-noise ratio. This signal-to-noise ratio is called *effective* because it may be determined by several factors. The most obvious one is background noise, which contributes directly to the signal-to-noise ratio. However, products of distortions in the time domain and non-linearity's are also considered as noise. This is derived by the specific design of the test signal. In Fig. 3.1 an illustration is given of the estimation of the signal-to-noise ratio within each frequency band. The test signal consists of a noise signal with a frequency spectrum equal to the long-term frequency spectrum of the speech signal. Each octave-band is modulated with a periodic signal in such a way that the *intensity envelope*² is modulated sinusoidal. This is indicated in Fig. 3.1 for the octave band with centre frequency 250 Hz. The modulation index (m) in this example is $m = 1$ at the input side and reduced to $m = 0.5$ at the output side.

2) The addition of uncorrelated signals (echoes, reverberation, and masking noises) is based on intensity summation. For instance, the addition of two sinusoidal modulated signals (same modulation frequency) with uncorrelated carriers will consist of a signal with a sinusoidal envelope modulation being the vector summation of the sinusoidal envelope of the two primary signals. This statement is only valid for intensity modulations, and not for amplitude modulations.

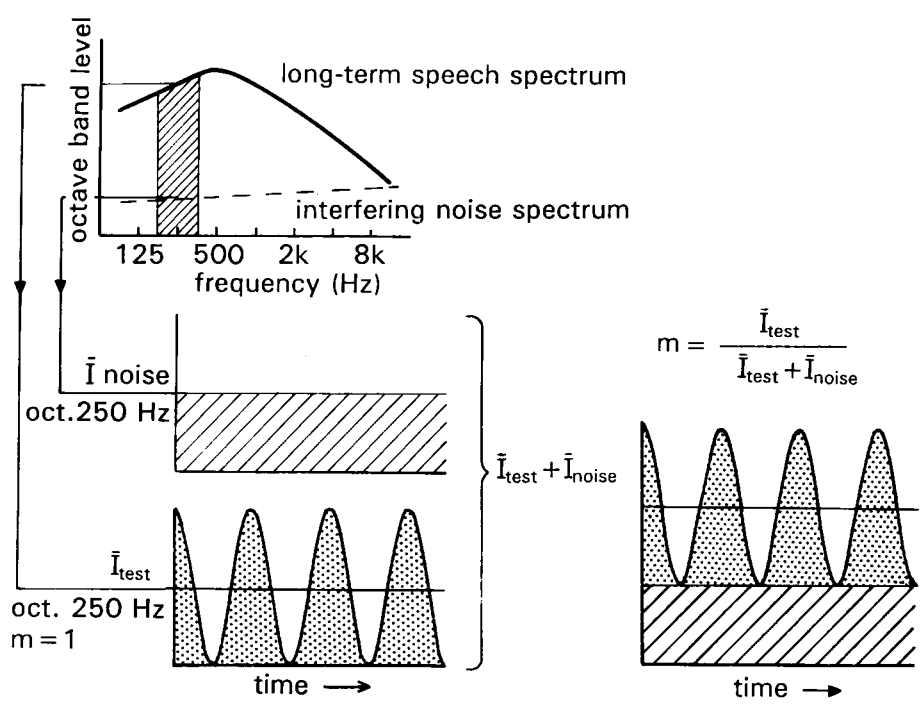


Fig. 3.1 Illustration of the effect of interfering noise on the modulation index m of a test signal.

Noise may be added to the test signal and the resulting envelope is obtained by addition of the intensity of both signal envelopes. Hence, the resulting envelope of this example is defined by a steady noise envelope (of a stationary noise signal) and the test-signal envelope. The resulting modulation index (m), being the test-signal intensity divided by the total intensity (test signal and noise), is directly related to the signal-to-noise ratio (SNR)

$$\text{SNR} = 10 \log \frac{m}{1-m} \text{ dB} \quad (1)$$

according to:

As described in chapter 2, the envelope function of a fluctuating speech signal contains a range of frequencies, representing the succession of speech events from the shortest speech items (such as plosives) up to words and sentences. Due to distortion in the time domain (reverberation, echoes, and automatic gain control) this fluctuation pattern may be affected, in this way reducing intelligibility. This is modelled in the STI procedure by determining the modulation transfer function for the range of relevant frequencies present in the envelope of *natural* speech signals. As described before (Steeneken and Houtgast, 1980) a relevant range for these modulation frequencies extends from 0.63 Hz up to 12.5 Hz. Separation in 1/3-octave steps, yields 14 bands. This results in a measuring procedure

according to Fig. 3.2 where the modulation transfer index, m , for each octave band (125 Hz - 8 kHz) and each modulation frequency (0.63 - 12.5 Hz) is determined separately. The figure gives the measuring set-up for one octave band. A noise signal with the required frequency spectrum (normally the long-term speech spectrum) is amplitude modulated by a signal $\sqrt{1 + \cos(2\pi \cdot f_m \cdot t)}$ which results in a sinusoidal intensity modulation $I \cdot \{1 + \cos(2\pi \cdot f_m \cdot t)\}$. This modulation function can be obtained digitally and can be generated by computer. At the receiving side, octave-band filtering and (intensity) envelope detection is applied. From the resulting envelope function a Fourier analysis determines the modulation index reduction, due to the reduction by the transmission channel. This procedure is repeated for each cell of the matrix given in Fig. 3.2. It should be noted that the block diagram of Fig. 3.2 represents only one channel corresponding with one octave band. The original set-up consists of a set of separate channels for all octave bands considered.

With the test signal as described above, distortions such as band-pass limiting, and noise masking, as well as distortion in the time domain can be dealt with. Non-linear distortions, however, have to be modelled additionally. If a speech signal is passed through a system with a non-linear transfer (e.g. peak clipping or quantisation), harmonic distortion components and inter-modulation components will be produced in other frequency bands. For this reason the test signal should not be modulated with one and the same modulation frequency for all octave bands simultaneously. Otherwise, non-linear distortion components cannot be discriminated from the modulated test signal in the frequency band considered. Therefore, in the case of non-linear distortion, all frequency bands, except the one under test, are modulated with uncorrelated signals so that the envelopes of the distortion components are not correlated with the test-signal envelope in the octave band under test. Such distortion components are then considered as noise (they add to the noise in the octave band under test) and reduce the effective signal-to-noise ratio in a similar way as would occur with other interfering signals. The relative levels of the test signal in the octave bands with the uncorrelated (speech-like) envelope were adjusted for optimal prediction of intelligibility in non-linear transfer conditions. The consequence of this procedure is a successive measurement for each of the seven octave bands rather than a simultaneous measurement as can be applied for communication channels with a linear transfer.

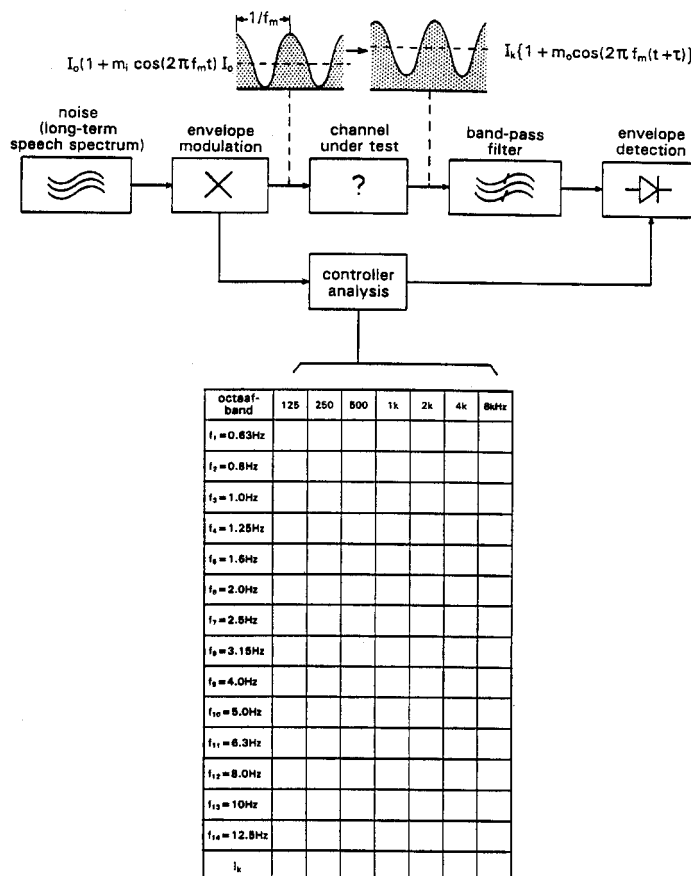


Fig. 3.2 General block diagram of the measuring set-up. The modulation index reduction at the output (m) is determined for all cells of the matrix (7 octave bands and 14 modulation frequencies). Also the octave levels (I_k) are obtained, for calculation of the auditory spread of masking.

Besides the masking introduced by the noise in the transmission channel two other factors have to be taken into account: (1) an additional auditory masking phenomenon³ (auditory spread of masking) and (2) the absolute hearing threshold. Both effects are modelled as an imaginary masking noise that leads to a decrease of the effective signal-to-noise ratio. Hence, resulting in a reduction of the modulation transfer index m . For this purpose not

- 3) Auditory spread of masking is the effect, introduced by the hearing organ, that a strong masker in a lower frequency range may reduce the perception of a tone or narrow-band signal. The amount of masking depends on the level difference between masker and masked signal, on the absolute level of the masker, and on their frequency distance. Zwicker and Feldtkeller (1967) give a detailed description.

only the modulation transfer has to be determined but also the signal levels in the frequency bands have to be considered. In Fig. 3.3 the effect of the masking by frequency band (k-1) upon frequency band k is indicated for a signal level of 60 dB SPL. The masking as a function of the signal level is given in Table 3.1.

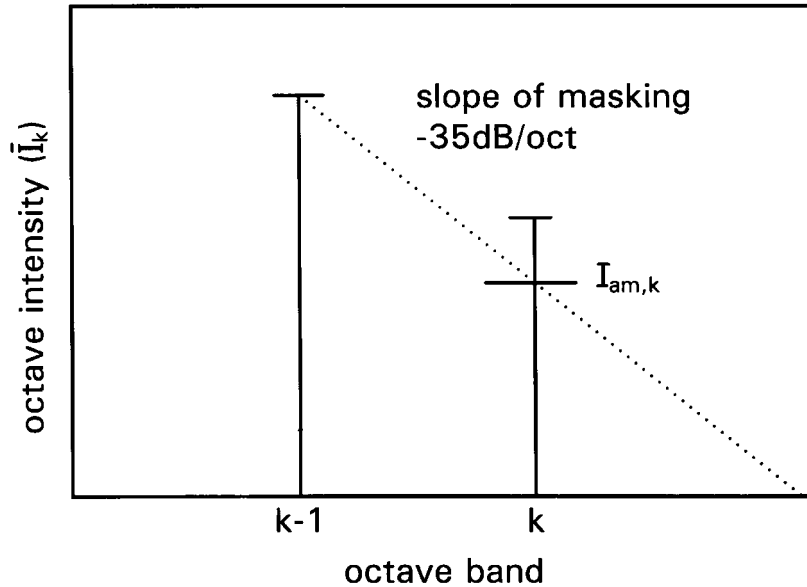


Fig. 3.3 Auditory masking of octave band k-1 upon the next higher octave band k. The slope of the masking effect versus frequency band corresponds to -35 dB/oct. This is equivalent to an auditory masking factor of $\text{amf} = 0.000316$.

The masking effect, as modelled in the STI approach, does not depend on the frequency band considered but does depend on the level. For example, the slope of masking decreases with 35 dB/oct for signal levels between 55 and 65 dB. The corresponding auditory masking factor (amf) of the intensity of the primary masking signal amounts $\text{amf} = 0.000316$ (intensity attenuation of masking signal upon adjacent next higher octave band). As the masking effect by only one lower frequency band is considered, the intensity of the

$$I_{\text{am},k} = I_{k-1} * \text{amf} \quad (2)$$

masking signal becomes:

where $I_{\text{am},k}$ represents the intensity level of the auditory masking signal for octave band k, and I_{k-1} represents the signal intensity of octave band (k-1).

In Table 3.1 the slope of the masking as a function of the octave level is given.

Table 3.1 Octave level specific slope of masking

Octave level dB	46-55	56-65	66-75	76-85	86-95	>95
Slope of masking	-40	-35	-25	-20	-15	-10
Auditory masking factor	0.000100	0.000316	0.003162	0.010000	0.031622	0.100000

The effect of the absolute hearing threshold is modelled in the STI-approach as the lower limit of the masking noise level within each octave band ($I_{rs,k}$, see Table 3.2). This level is only relevant if I_k refers to the presentation level to the listeners.

The auditory spread of masking and the hearing threshold are accounted for by a reduction in the modulation index. The corrected modulation index becomes:

$$m'_{k,f} = m_{k,f} \frac{I_k}{I_k + I_{am,k} + I_{rs,k}} \quad (3)$$

where $m_{k,f}$ represents the modulation index for octave band k and modulation frequency f , and m' the corrected modulation index.

The effective signal-to-noise ratio for octave band k and modulation frequency f then becomes:

$$SNR_{k,f} = 10 \log \frac{m'_{k,f}}{1 - m'_{k,f}} \text{ dB} \quad (4)$$

According to the STI concept a signal-to-noise ratio between -15 dB and 15 dB is linearly related to a contribution to intelligibility of between 0 and 1. Therefore, the effective signal-to-noise ratio is converted to transmission index ($TI_{k,f}$), specific for octave band (k) and modulation frequency (f), by the equation:

$$TI_{k,f} = \frac{SNR_{k,f} + \text{shift}}{\text{range}}, \quad \text{where } 0 \leq TI_{k,f} \leq 1.0. \quad (5)$$

The shift equals 15 dB and the range equals 30 dB. In this way a relation between the effective signal-to-noise ratio and the TI is obtained as shown in Fig. 3.4.

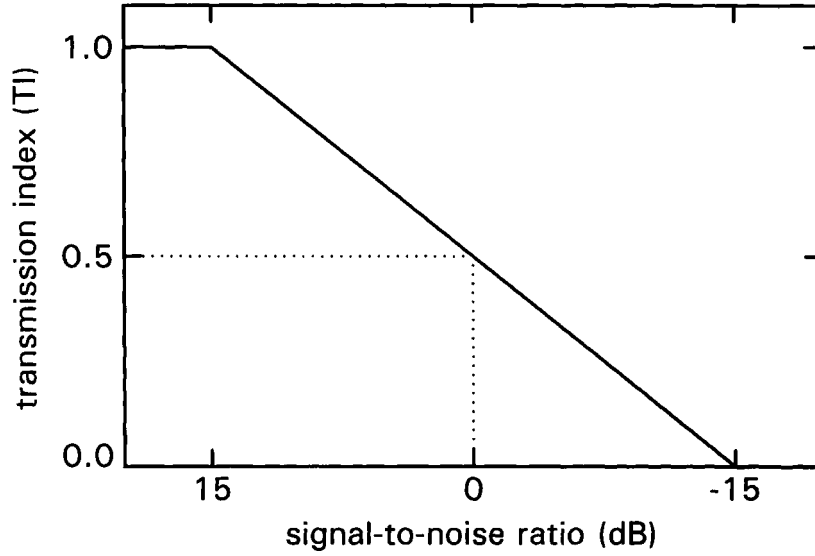


Fig. 3.4 Relation between the effective signal-to-noise ratio and the transmission index for a shift of 15 dB and a range of 30 dB.

All 14 transmission indices related to modulation frequencies between 0.63 and 12.5 Hz⁴, are obtained for each octave band. The mean of these indices results in the modulation transfer index (MTI_k) and is specific for the contribution of octave band k. The MTI_k is given by:

$$MTI_k = \frac{1}{14} \sum_{f=1}^{14} TI_{k,f} \quad (6)$$

Finally, according to the revised formula, the STI_r is obtained by a weighted summation of the modulation transfer indices for all seven octave bands and the corresponding redundancy correction. This is given by:

$$STI_r = \alpha_1 \cdot MTI_1 - \beta_1 \cdot \sqrt{(MTI_1 \cdot MTI_2)} + \alpha_2 \cdot MTI_2 - \beta_2 \cdot \sqrt{(MTI_2 \cdot MTI_3)} + \dots + \alpha_7 \cdot MTI_7 \quad (7)$$

4) This range provides an optimal fit for conditions with temporal distortions in relation to conditions with noise distortion.

where,

$$\sum_{k=1}^7 \alpha_k - \sum_{k=1}^6 \beta_k = 1. \quad (8)$$

The factor α_k represents the octave-weighting factor and β_k the so-called redundancy correction factor. This redundancy correction is related to the contribution of adjacent frequency bands. Steeneken and Houtgast (1999, 2002a, 2002b) describe the optimal weighting factors and redundancy factors for male and female speech and different groups of phonemes.

In Table 3.2 the α and β values are given for male and female speech, also the level of the reception threshold (eq. 3) is given in decibel.

A flow diagram of the calculation procedure of the STI is given in Fig. 3.5.

Table 3.2. STI_r octave-band specific male and female weighting factors and the absolute reception threshold in decibel.

Octave band Hz		125	250	500	1k	2k	4k	8k
Males	α	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	β	0,085	0,078	0,065	0,011	0,047	0,095	–
Females	α	–	0,117	0,223	0,216	0,328	0,250	0,194
	β	–	0,099	0,066	0,062	0,025	0,076	–
Absolute reception threshold dB	$L_{rs,k}$	46	27	12	6,5	7,5	8	12

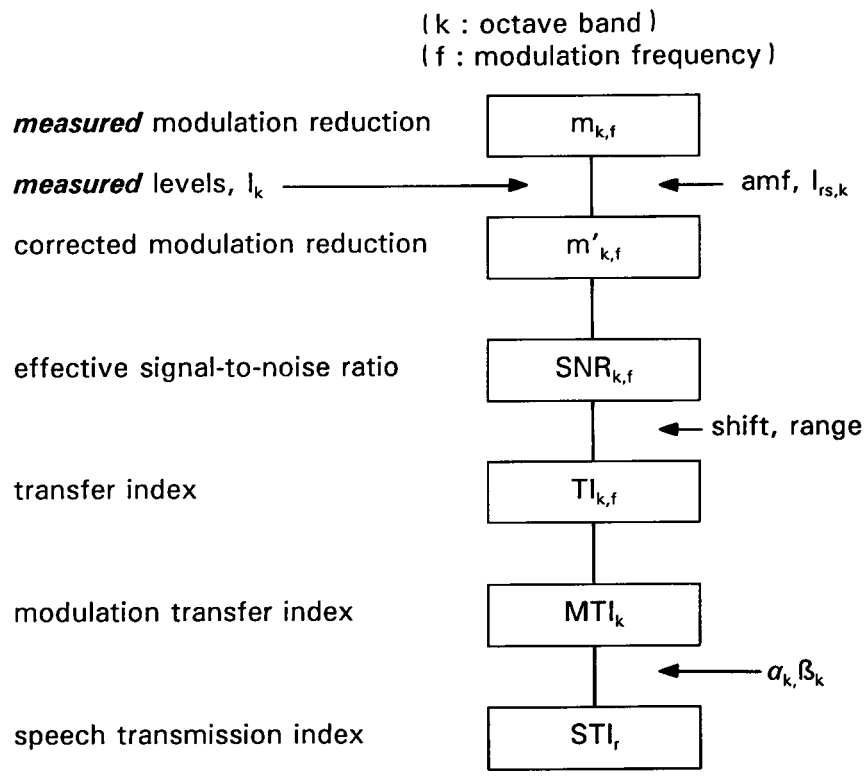


Fig. 3.5 Flow diagram of the STI calculation scheme.

Some simplifications of the procedure described above were made in order to decrease the measuring time, but these simplifications restrict the range of applicability. The measurement of a complete matrix of 98 m-values according to Fig. 3.2 and a measuring time for each m-value of 10 s results in a total measuring time of 15 minutes.

A reduction of the 14 modulation frequencies to only three modulation frequencies results in a total measuring time of less than 4 minutes, but as a consequence no complete modulation transfer is obtained. This means that distortions in the time domain are not accounted for correctly. Therefore, this method is normally used only for communication channels with no degradation due to echoes or reverberation such as with person-to-person communication.

Also, the number of octave bands considered may be reduced. This is the case with the RASTI method, where only the contributions of the modulation transfer for the octave bands with centre frequencies 500 Hz and 2 kHz are considered. This can be used as a screening approach for direct person-to-person applications.

Another simplification can be applied to the test signal if the uncorrelated (speech-like) modulations, required for the correct interpretation of non-linear distortions, are omitted. This opens the possibility of applying a simultaneous modulation and parallel processing of

all frequency bands, thus decreasing the measuring time. This procedure is used in the STITEL and STIPA method that requires a measuring time of about 15 s.

It should be noted that the STI method can be applied to transmission channels with the type of distortions listed before. Due to the specific compilation of the test signals and the type of analysis some types of distortions are not accounted for. These are:

- frequency shifts (such as obtained with single side-band transmission),
- frequency multiplication (such as obtained with analogue tape recorders which run at an incorrect tape-speed), and
- vocoders (systems which introduce errors related to voiced-unvoiced speech fragments and pitch errors).

4 Overview methods, test signals, and calculation constants

The 'full' STI method includes measurements within seven octave bands and 14 modulation frequencies within each octave band. However, certain applications do not require such a robust measuring scheme. For those measurements specific simplifications of the measuring method can be applied in order to increase the measuring efficiency. The various simplifications of the measuring procedure have led to different measuring schemes that are adapted for specific groups of applications

Respective versions are:

STI-14: A universal measuring scheme, which is applicable to all types of communication systems (except vocoders), includes a successive measurement of the full matrix as given in Fig. 3.2. This method is called STI-14 and refers to the original. For this method test signals for seven octave bands and 14 modulation frequencies are transmitted and analysed successively.

STI-3: As the STI-14 method is time consuming a limitation in the modulation frequency domain is applied in order to decrease the measuring time. This version, based on three modulation frequencies, has limited applicability with respect to conditions with distortions in the time domain (the resolution is decreased). The measuring method is referred to as STI-3.

STITEL: The STITEL (Speech Transmission Index for TELEcommunication channels) is a stripped version of the STI and has no robust coverage for transmission channels with distortion in the time domain and for non-linear systems.

STIPA: The STIPA (Speech Transmission Index for Public Address systems) is a stripped version of the STI-14 and has a robust coverage for distortions in the time domain and limitations in the frequency domain. A limited coverage of non-linear distortions is obtained.

RASTI: The RASTI system (Room Acoustical⁵ Speech Transmission Index) is based on the MTF for only two octave bands, no coverage for band-pass limiting and non-contiguous noise spectra is obtained. This method is developed for person-to-person communications in a room acoustical environment and does account for distortion in the time domain.

An overview of these methods is given in Table 4.1. The field of application is also indicated. For some programs the applicability is condition dependent (e.g. the type of non-

5) Sometimes referred to as RApid Speech Transmission Index.

linear distortion or the type of reverberation). This means that a test with the STI-14 or STI-3 has to be performed in order to verify the applicability.

Table 4.1 Overview of the measuring procedures, the applications, and the corresponding test signals.

Application	Band-pass limiting	Non Linear Distortion	Reverberation Echoes	Test signal types	Measuring time
STI-14 (7 octaves, 14 fmod)	Yes	Yes	Yes	Male, female	15 min
STI-3 (7 octaves, 3 fmod)	yes	Yes	Condition Dependent	Male, female	4 min
STITEL (7 octaves, 7 oct. related fmod)	yes	Condition Dependent	Condition dependent	Male, female, Original, Phoneme groups	15 s
STIPA 7 octaves, 14 oct. related fmod)	yes	Condition Dependent	Yes	Male, female	15 s
RASTI (2 octaves, 4-5 fmod)	no	No	Yes	Original	15 s

The frequency weighting and redundancy correction factors are identical for the STI-14, STI-3, STITEL and STIPA method but different for male and female speech. For the RASTI only two octave bands are used (500 Hz and 2 kHz).

5 Interpretation of the STI value: relation with subjective measures

The use of the STI-method for more than 30 years, the international application, and the validation in other studies (Houtgast and Steeneken, 1984; Anderson and Kalb, 1987; Barnett, 1999; Mapp, 2001; van Wijngaarden and Steeneken, 1999) has led to a robust qualification of the STI value in terms of speech intelligibility. The validation of the method

with different intelligibility tests resulted into a robust relation with a variety of subjective measures. In Fig. 5.1 this relation for the original STI concept and various intelligibility measures is given. It should be noted that the earlier experiments were designed to establish the optimal relation for CVC words of the type “phonetically balanced” for Dutch nonsense words. In later studies CVC words with a uniform phoneme distribution were used. This introduced a slightly different relation between STI and CVC-word score. All the data in this manual refer to CVC-word lists with such uniform (equally balanced) phoneme distribution and nonsense words.

The improvement of the STI method by the introduction of the redundancy corrections resulted in essentially the same relation between the CVC-word score and the STI. However, the STI values obtained according to the new method are referred to as STI_r . The improvement becomes apparent mainly when transmission channels with severe band-pass limitation, non-contiguous frequency transfer or masking noise with a discontinuous spectrum are tested. In Fig. 5.1 the relation between the STI_r , the CVC-word score, and sentence intelligibility (short simple sentences) is given for male speech. Additionally the relation between the STI_r and the CVC-word score for female speech is given in Fig. 5.3.

The relation between the STI_r , the CVC-word score and phoneme group scores can also be derived from the expressions given in Table 5.1.

$$\text{predicted score} = \{A * e^{(B*STI)} + C\} * 100 \quad (\%) \quad (9)$$

Table 5.1 Relation between the STI_r , the CVC-word score, and phoneme-group scores for male and female speech.

Word or phoneme type	Male			Female		
	A	B	C	A	B	C
CVC words	-1.5301	-2.0	1.15	-1.7584	-1.5	1.37
Fricatives	-0.9000	-4.2	0.90	-0.9466	-4.1	0.90
Plosives	-1.1531	-4.1	1.01	-1.1256	-6.0	0.95
Vowel-like consonants	-1.4602	-4.2	1.05	-1.3216	-4.0	1.09
Vowels	-0.9976	-2.9	1.03	-1.2057	-3.1	1.04

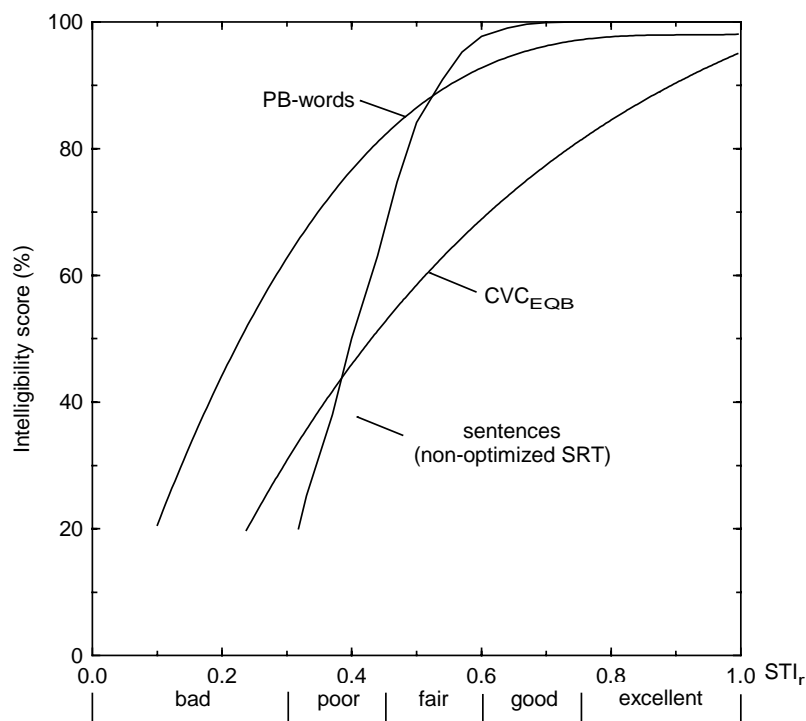


Fig. 5.1 Qualification of the STI_r (Steeneken and Houtgast, 2002b) and relation with various subjective intelligibility measures for MALE speech.

As indicated before, the STI method can also be used to predict the intelligibility scores for certain phoneme groups. For this purpose specific test signals (corresponding to the mean phoneme-group spectrum) and frequency weightings and redundancy correction factors are used. It should be noted that this method couldn't be used for all types of channels. Specifically channels with a "memory" (such as reverberation or automatic gain control) are affected by the level of embedded signals that may interact with the various (phoneme-group-specific) test-signal levels. The relation between the phoneme-group specific STI (referred to as STI_s) and the phoneme-group score is given in Figs 5.2 and 5.3 (for male and female speech, respectively). The equations for calculating the various phoneme-group scores are given in Table 5.1.

Besides a direct estimation of the CVC-word score this score can also be predicted by combining phoneme-group scores obtained from the STI_s values for the fricatives, plosives, vowel-like consonants, and vowels. This is performed in two steps: (1) calculation of the initial consonant and final consonant score (a weighted combination of the plosive, fricative, and vowel-like consonants scores), and (2) calculation of the CVC-word score from the product of the initial consonant, vowel, and final consonant probabilities.

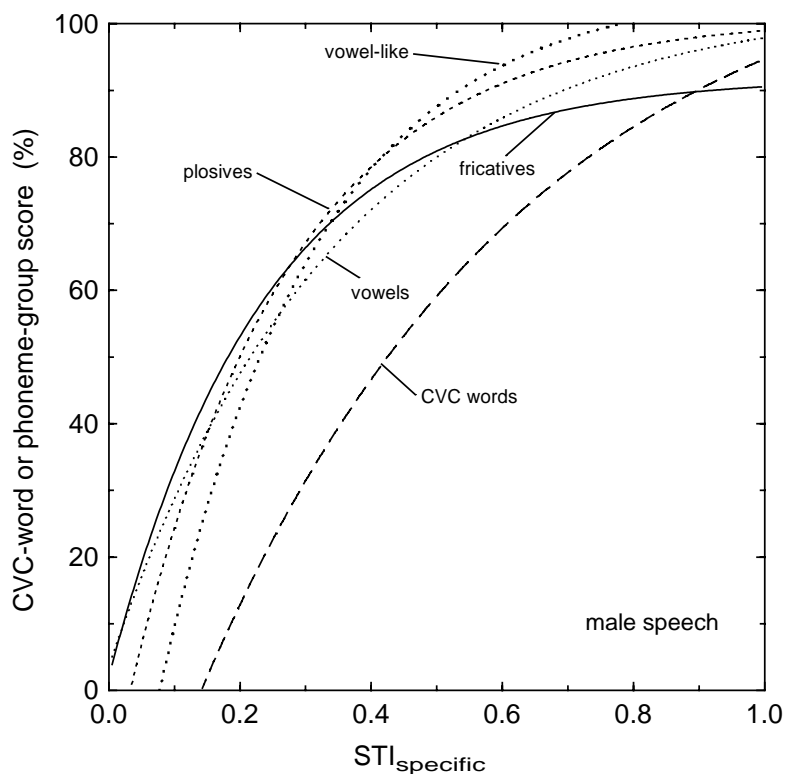


Fig. 5.2 Relation between predicted phoneme-group scores and the corresponding phoneme-group-specific STI_s for MALE speech. The relation for the CVC-word score is also given.

The advantage of predicting the word score by a (weighted) combination of the predicted phoneme-group scores is that it is not restricted to the example with the equally balanced CVC words, but that it can also be used to predict the word score of PB-words or any other combination. The restriction is, however, that the word score is indeed defined by independent phoneme scores.

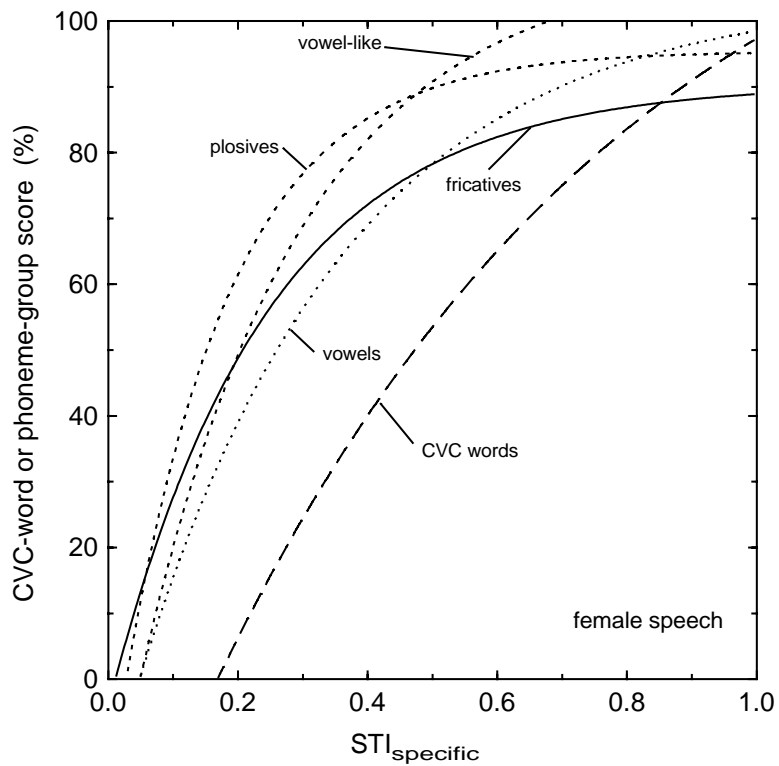


Fig. 5.3 Relation between predicted phoneme-group scores and the corresponding phoneme-group-specific STI_s for FEMALE speech. The relation for the CVC-word score is also given.

The relative test signal spectra for phoneme groups and the embedded CVC test words are given for males and females in Figs. 5.4 and 5.5.

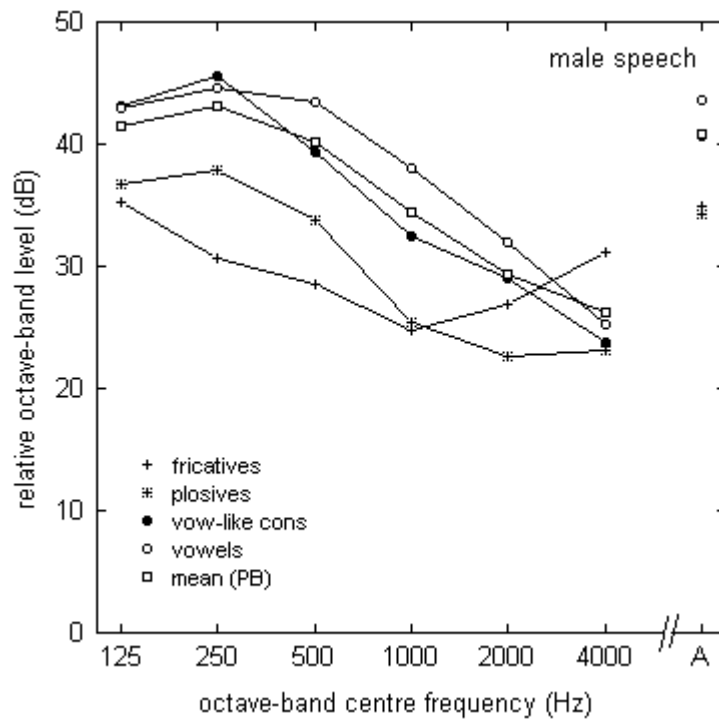


Fig. 5.4 Relative test signal spectra for the four phoneme groups and for phonetically balanced speech (connected discourse). The dBA values represent the relative level of each group for connected discourse of MALES.

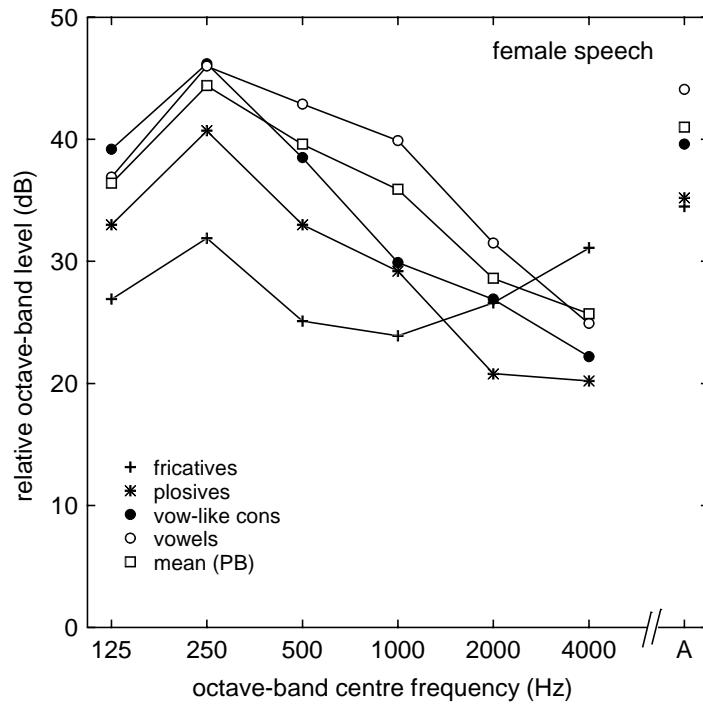


Fig. 5.5 Relative test signal spectra for the four phoneme groups and for phonetically balanced speech (connected discourse). The dBA values represent the relative level of each group for connected discourse of FEMALES.

6 Diagnostic features, some examples

The STI-method allows for two types of diagnostic analysis (1) based on the analysis of the test signal, and (2) based on the type and level of the test signal.

(1) As shown in Fig. 3.2 the modulation index reduction (effective signal-to-noise ratio) is obtained for seven octave bands and 14 modulation frequencies. The contribution of each octave band to the STI-value represents information on the frequency response of the system and on the spectrum of a masking signal. Generally, a low modulation index in combination with a low octave level indicates a poor frequency response. However, a low modulation index in combination with a high octave level represents a high impact of a masking signal. In Table 6.1 an example is given of a communication system with a limited frequency transfer. The table represents a typical output of the STI-calculation using the STITEL program. Both the STI_r (according to the concept including a redundancy correction) and the STI (according to the concept given by Steeneken and Houtgast, 1980) are given. Also the CVC-word score, based on the STI_r value and on the test signal type (male, female) are presented. The signal level at the input of the analogue-to-digital system is given and expressed in $dB\mu V$ (if no additional calibration correction is applied). This example refers to the frequency transfer of a normal telephone channel. By comparison of the spectrum of the input signal and the output signal the frequency transfer can be obtained. This method is only valid if no additional noise is added between input and output. As mentioned above this can be detected by the reduction of the modulation transfer (all the TI's are close to '1').

Table 6.1 Example of the STI value, levels, and octave-band specific information for a transmission channel with a limited frequency transfer obtained with the STITEL method.

STI_r	=	0.89	(Male speech, corresponding CVC-word score 89%)						
STI	=	0.86							
Level	=	110.0 dB	Level correction = 0.0 dB						
Level (A)	=	107.8 dBA	AD/DA range : 6.0 V(pp) equals 16 bit						
Octave centre freq.		125	250	500	1000	2000	4000	8000	Hz
Octave level		70.5	101.8	107.4	103.6	98.4	82.1	52.1	dB
Mod. Index (m)		0.97	1.00	1.03	1.02	1.00	1.01	0.01	
m-correction		1.00	1.00	1.00	1.00	1.00	0.99	0.76	
Transm. Index (TI)		1.00	1.00	1.00	1.00	1.00	1.00	0.00	
Relative Freq-resp.		-40.22	-8.86	0.39	2.63	3.45	-6.85	-30.88	dB
Modulation Frequency		1.12	11.33	0.71	2.83	6.97	1.78	4.53	Hz

Table 6.2 Example of the STI value, levels, and octave-band specific information for a transmission channel with a limited frequency transfer and a white noise masking signal (signal-to-noise ratio 0 dBA).

STI _r	=	0.48	(Male speech, corresponding CVC-word score 56%)						
STI	=	0.48							
Level	=	108.5 dB	Level correction = 0.0 Db						
Level (A)	=	107.3 dBA	AD/DA range : 6.0 V(pp) equals 16 bit						
Octave centre freq.		125	250	500	1000	2000	4000	8000	Hz
Octave level		67.7	99.1	104.8	102.1	101.0	96.3	53.6	dB
Mod. Index (m)		0.92	0.92	0.95	0.72	0.27	0.02	0.02	
m-correction		1.00	1.00	1.00	1.00	1.00	1.00	0.15	
Transm. Index (TI)		0.86	0.85	0.92	0.64	0.36	0.00	0.00	
Relative Freq-resp.		-42.53	-11.04	-1.72	1.66	6.52	7.82	-28.88	dB
Modulation Frequency		1.12	11.33	0.71	2.83	6.97	1.78	4.53	Hz

An example of a combination of band-pass limiting and additive noise is given in Table 6.2. As the noise signal used for this example is white noise (increase of 3 dB per octave band) the modulation indices for the higher octaves are lower than those for the low frequency bands (the TI's decrease from .092 to 0.02).

The modulation transfer function (MTF) offers information concerning the type of distortion in the time domain. If only a stationary noise is added to the speech or test signal, the decrease of the modulation transfer will be modulation-frequency independent. This is illustrated in Fig. 6.1. The reduction of the modulation transfer (m) is also given as a function of the signal-to-noise ratio.

In the case of distortion in the time domain (automatic gain control, echoes, and reverberation) a modulation-frequency specific reduction will be obtained. Reverberation acts as a low-pass filter on the fluctuations of the envelope. This is shown by the MTF given in Fig. 6.2. In this graph also the theoretical relation between the modulation reduction (m) and the reverberation time (T) is given according to Houtgast and Steeneken (1973).

For echoes a rippled modulation transfer is obtained. For a fixed echo delay time (τ) the modulated envelope of the reflected signal (relative level δ) will, as a function of the modulation frequency, vary in phase with respect to the primary signal. This result in a rippled modulation transfer function (MTF). In Fig. 6.3 an example of such a MTF is given. The theoretical relation is also given.

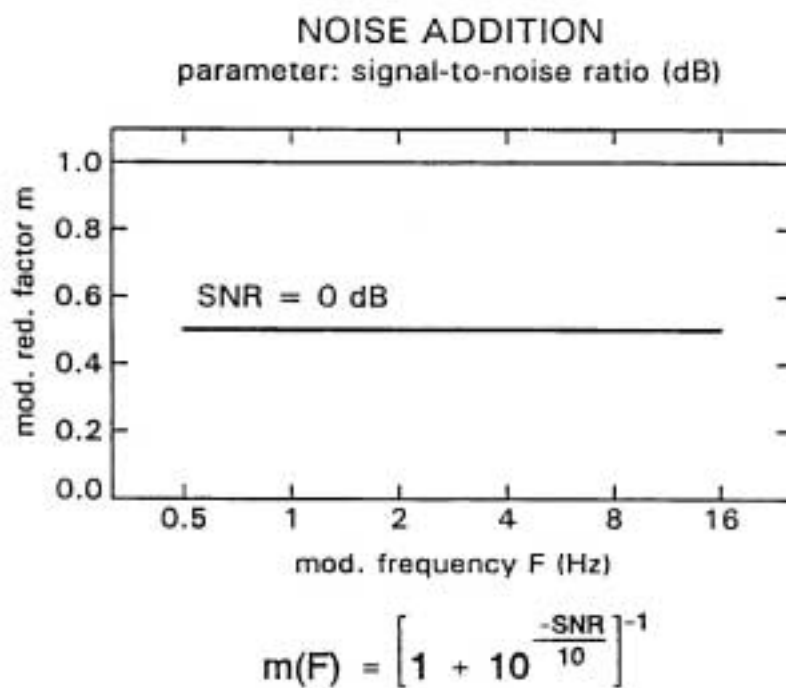


Fig. 6.1 Example of the modulation transfer function for conditions with noise.

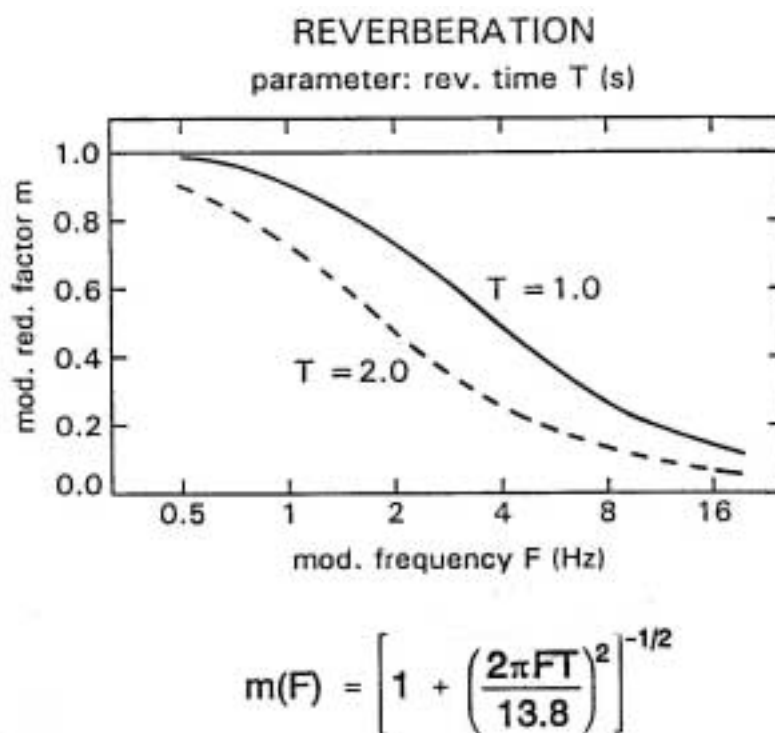


Fig. 6.2 Example of the modulation transfer function for conditions with reverberation.

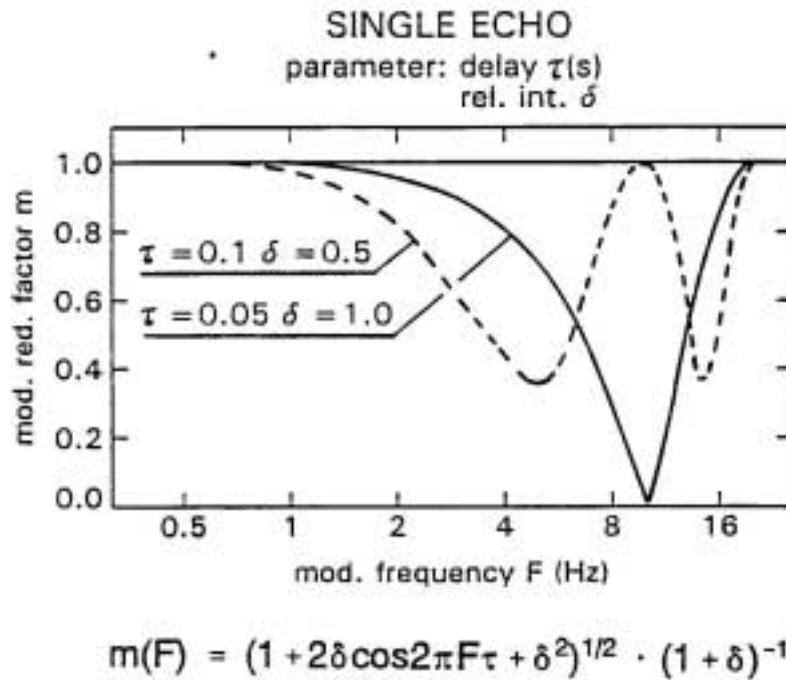


Fig. 6.3 Example of the modulation transfer function for conditions with echoes.

Automatic gain control systems mainly reduce slow level variations, which may be described as a high-pass filter applied to the envelope fluctuations. Here, the relation between the modulation reduction and the attack and release time of the AGC is complex and cannot be represented by a simple formula. It is obvious that for systems with a distortion in the time domain the full MTF should be determined, hence based on the complete matrix of Fig. 3.2.

(2) Another method to obtain diagnostic information is to vary the test signal level or the type of test signal. Variation of the test-signal level will discriminate between signal-level dependent and signal-level independent distortions. For example, the effect of masking noise will increase at lower test signal levels while the effect of reverberation and echoes is not signal-level dependent. This feature is often used in the evaluations in room acoustics and becomes even more powerful if it is used in combination with the frequency dependent analysis as mentioned above.

As described in section 3.1, a specific test signal is applied for non-linear communication channels. While performing an analysis in one of the seven octave bands, uncorrelated fluctuations are present in the other six octave bands. These may introduce distortion components within the octave band under test and hence reduce the modulation transfer. This is similar as with speech signals. Comparison of the modulation transfer (or the STI) for a given channel measured with two types of test signals, one with the representative uncorrelated fluctuations present and one measured without these fluctuations, will show the effect of the deterioration by the non linear frequency transfer.

7 Speech and test-signal level adjustment

For reproducible experiments concerning the effect of noise on speech transmission quality, it is important to specify the speech levels, the noise levels and the corresponding signal-to-noise ratios.

Various studies (Brady, 1965; Kryter, 1970; Berry 1971; Steeneken and Houtgast, 1978, 1986) have defined speech level measures. It was also shown that a signal-to-noise ratio variation of only 1-2 dB might have the same effect on the results as typical speaker and inter-listener variations. We therefore specified a method for measuring speech levels and noise levels, which offers such a resolution. The measure should be robust for the various speech types (male/female, connected discourse/isolated words), recording conditions (background noise, frequency transfer), and should also be applicable to noise signals. We have developed such a measure (Steeneken and Houtgast, 1978, 1986) mainly for adjusting the signal level of the STI test signal to the speech level for similar conditions. The measuring method was made generally available by development of a, platform independent, digital signal-processing algorithm.

7.1 Speech level measuring method

A high correlation was found between the speech level and the speech intelligibility for level measures based on frequency-weighted speech signals with a reduced contribution of frequency components below approx. 250 Hz (Kryter, 1970; Steeneken and Houtgast, 1978, 1986). The standardised frequency-weighting function according to the A-filter was used for this purpose (standardised for acoustical measurements).

After filtering, the running (intensity) envelope is determined by squaring and low-pass filtering (47 Hz) the waveform. From this envelope function the envelope distribution histogram is obtained, and the RMS value can be computed from this histogram. The advantage is that the RMS value can also be obtained for values above a certain level after sampling. In order to compare the level of short speech tokens (simple words altered with long silent periods) and the level of connected discourse, a level threshold for suppression of the silent periods is required. Hence, this threshold is applied to the envelope function of the speech signal rather than to the waveform, and therefore does not affect each zero crossing of the speech signal. The threshold level is defined to be 14 dB below the resulting RMS level (iterative procedure). This definition is signal-related and does not strongly depend on other effects such as background noise level (down to signal-to-noise ratios of 4 dB), shape of the envelope distribution, etc. The same principle can be applied to stationary noises but in that case the threshold function is not effective.

The relation between various level measures obtained from two types of speech signals (connected discourse, and CVC words in a short carrier phrase) is given in Fig. 7.1.

The level measures are: the 1% peak level (1% overflow criterion), the mean of the peak deflections of a sound level meter set to "fast" (dBA fast), the RMS values obtained with a

8 Application examples

The STI method can be used for speech communication systems: (1) radio links, intercoms or digital (waveform based) speech coders, (2) electro acoustic transducers (microphone and telephone), and (3) for room acoustics. Although the STI measuring method for all three types of speech communication systems is the same, some system-specific simplifications of the measuring method are allowed. This leads to a faster result. Usually linear communication channels and electro-acoustic transducers (used close to the mouth or ear) can be assessed with the STITEL measuring program. For room acoustical applications the full STI measurement (STI-14, or STIPA) should be used. However for some specific applications (such as public address systems in open environment) we may also use STITEL. This has to be decided by making a reference measurement with STI-14 and check that echoes or reverberation does not affect the MTF.

The method of connecting the test signal to the system under test is different for the various systems. For communication systems an electrical input and output can be used. However, for applications with microphones or in room acoustics an artificial mouth has to be used in order to obtain an acoustically coupled test signal. For testing telephones and headsets an artificial ear is used. It is obvious that also a test of a complete communication system is possible (including microphone, communication system, headset and acoustically added background noise).

In the next sections some examples of these applications are given.

8.1 Communication channels

The first example concerns a diver underwater telephone system. The evaluation method of such a system with the STI-approach is similar to the method used for radio links or other transceivers. The effect of various parameters upon the transmission quality can be studied. The following parameters are of interest: the STI-value as a function of the range between transmitter and receiver, propagation conditions, and the input level of the modulator (especially if there is no automatic gain control).

The underwater telephone system presented in this example consists of a base station and a diver station. At the base-station side the acoustical transmitter receiver (a hydrophone) was placed in the water of a lake at a depth of 3 m. At various distances (4 m, 100 m, and 125 m) the diver set was put into the water at a depth of 3 m. Such an underwater telephone system is based on an amplitude modulated carrier with a carrier frequency between 8 and 40 kHz. This is similar to a radio-communication link but with a relatively low carrier frequency. The STI-test signal was electrically connected with the transmitter (base station). The test signal input level was variable. At the diver station side an electrical output (headphone) connection was used. In Fig. 8.1 the STI_r (obtained with the STITEL method) is given for the three distances between transmitter and receiver and as a function of the input level. For this type of application the maximum range is obtained at a STI_r of 0.35, which is related to a sentence intelligibility of just 100% (for very simple sentences). For

the 4 m and 100 m distance this STI_r value is obtained at various input levels. However at a distance of 125 m, which is a condition without a direct view between the two hydrophones, a very low STI_r is obtained.

In this example fixed conditions (distance between transmitter and receiver, and fixed input level of the modulator) were used. However, for some applications a continuously increasing range (e.g. a transmitter in a vehicle moving from or to the receiver) may be more appropriate. For this purpose a continuous analysis is made at the receiving side while at the transmitter side the test signal can be supplied from tape.

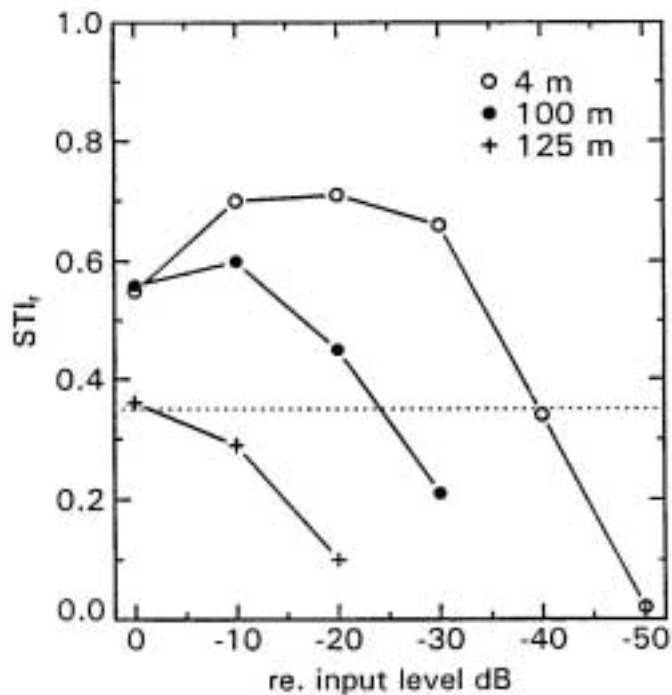


Fig. 8.1 STI_r as a function of the audio input level of the transmitter, for three distances between an underwater telephone base station and diver station at a carrier frequency of 40 kHz.

A second example of a communication channel concerns digital wave-form coders. With wave-form coders, parameters as bit-rate and bit errors are to be considered. We compared two CVSD systems (Continuous Variable Slope Delta Modulation) at a bit-rate of 8 kb/s and 16 kb/s. In the connection between the coder and decoder of the systems, random bit errors were introduced. The bit error rate could be varied in steps of 1%. In Fig. 8.2 the STI_r for both systems as a function of the bit error rate is given. The measurements were performed with the STI_3 -method making use of three modulation frequencies within each octave band and suitable for non linear distortion. The results show that system A offers a better performance than system B. It is also shown that system A gives the same intelligibility at 8 kb/s as system B at 16 kb/s. The results also show the robustness of these CVSD systems with respect to bit errors.

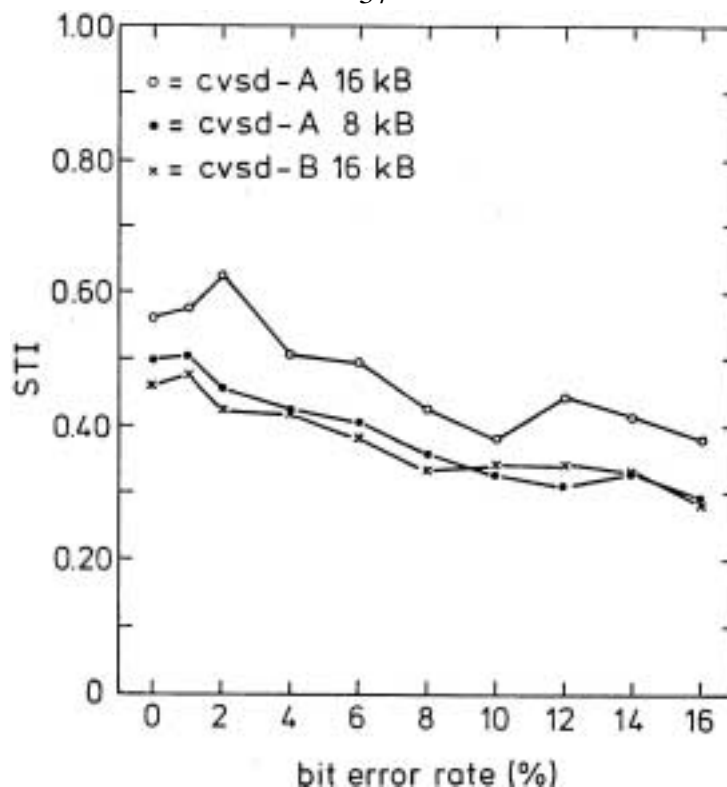


Fig. 8.2 STI_r for two CVSD systems, two bit rates as a function of the bit error rate.

8.2 Electro acoustic transducers

Microphones and telephones (headsets) are often used in noisy environments. Therefore, the assessment of these transducers should be performed in such an environment or by simulation. A second point of consideration for the use of a microphone is its position close to the mouth.

For the assessment of a microphone, an acoustical coupling is required. We developed an artificial mouth consisting of a (horn loudspeaker) driver unit, artificial head and connection tube between driver and outlet (mouth). The frequency transfer between the driver and the mouth is, due to the resonances in the tube, not flat. Therefore, the tube was filled with sound absorption material. This resulted in a frequency transfer which is flat within 10 dB. With the addition of a 1/3 octave equalizer a flat response between 100Hz and 10 kHz was obtained. The system was built into a box with the shape of a torso (see Fig. 8.3). At the moment of design no systems with suitable specifications were commercially available.



Fig. 8.3 Artificial mouth used for the assessment of microphones.

The level at 1 m distance in front of the mouth is typically 60 dBA. However to simulate a raised voice level (Lombard effect), the system can produce an undistorted signal with a level up to 75 dBA at 1 m distance. The radiation pattern is similar to that of humans. The system can also be used in room acoustics as an artificial speaker with a representative radiation.

Some artificial heads (including an artificial mouth and ears) are commercially available. It should be verified that the following specifications are fulfilled:

- (1) the frequency response must cover the frequency range of the STI-test signals (85 Hz - 11.2 kHz),
- (2) the maximum level at 1 m distance in front of the mouth must exceed 60 dBA, preferably 75 dBA,
- (3) the radiation pattern (also close to the mouth) must be representative for humans.

The artificial mouth, given in Fig. 8.3, is normally used in a high-noise room where a diffuse sound field can be produced. The microphone to be tested is placed at the required position in front of the artificial mouth. The STI is measured by connecting the test signal to the artificial mouth and by analysing the microphone output. The measurements are

normally performed at various microphone positions and various levels of the background noise.

In Fig. 8.4 the STI as a function of the noise level for two microphones is given.

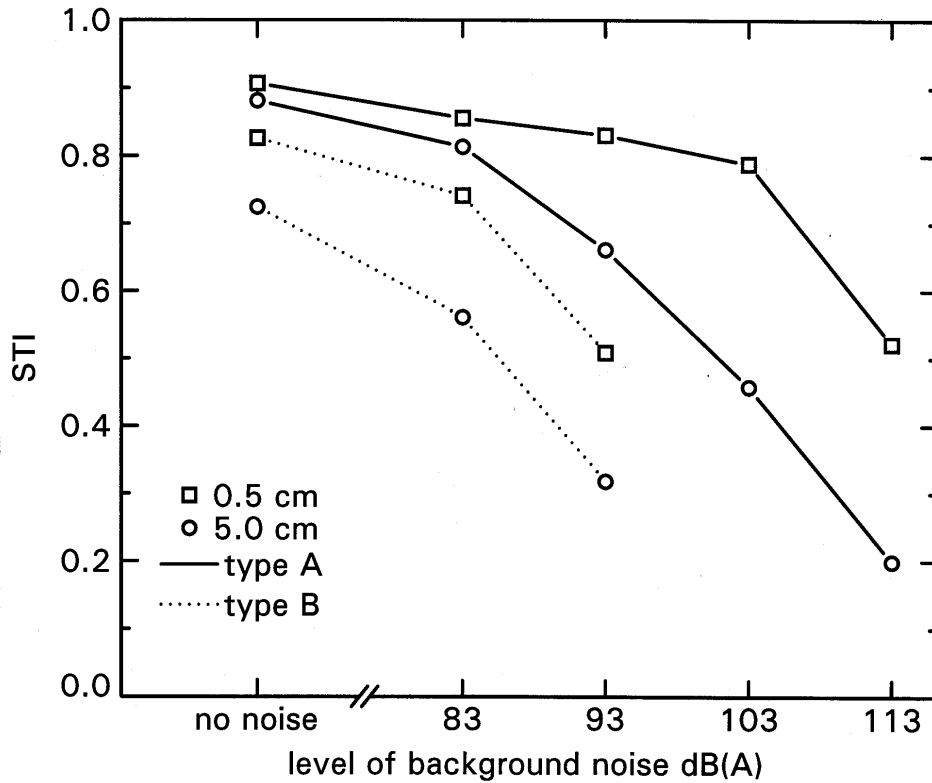


Fig. 8.4 STI_r for two microphones, at two positions in front of the mouth and as a function of the background noise level (for noise of a diesel engine).

For the assessment of telephones an artificial ear is required. Especially for the assessment of headphones mounted in earmuffs the head size, hair and wearing spectacles may influence the sound attenuation and the intelligibility. Therefore, normally a number of five subjects is used with a miniature electret microphone mounted near the ear canal. This is illustrated in Fig. 8.5B. The mounting and wiring of the microphone assembly is such that it does not interfere with the proper use of a telephone handset or a headset.

For measurements in combination with background noise a high-noise room with an adjustable noise level is used. The subject, with the (miniature) sense microphone mounted close to the ear-canal entrance, is positioned within this room. Special care must be taken that the subject is not exposed to sound levels above 85 dBA with unprotected ears. In order to obtain calibrated levels, the gain of the recording chain (microphone, microphone pre-amplifier and recording system) must be included in the STI measuring procedure (this can

be done by adjusting the correction factor in the configuration file of the STI-calculation program).

In general the presentation level of the speech (test) signal with a telephone is 60 - 75 dBA. Background noise levels may vary between 50 - 60 dBA (office) to 105 dBA (inside a fighter cockpit) or even up to 115 dBA (inside an armoured car or helicopter). In Fig 8.6 the STI_r is given for two types of telephone systems as a function of the background noise level (STITEL method).



Fig. 8.5 Subject positioned in a high-noise room and the mounting of the electret microphone near the entrance of the ear canal.

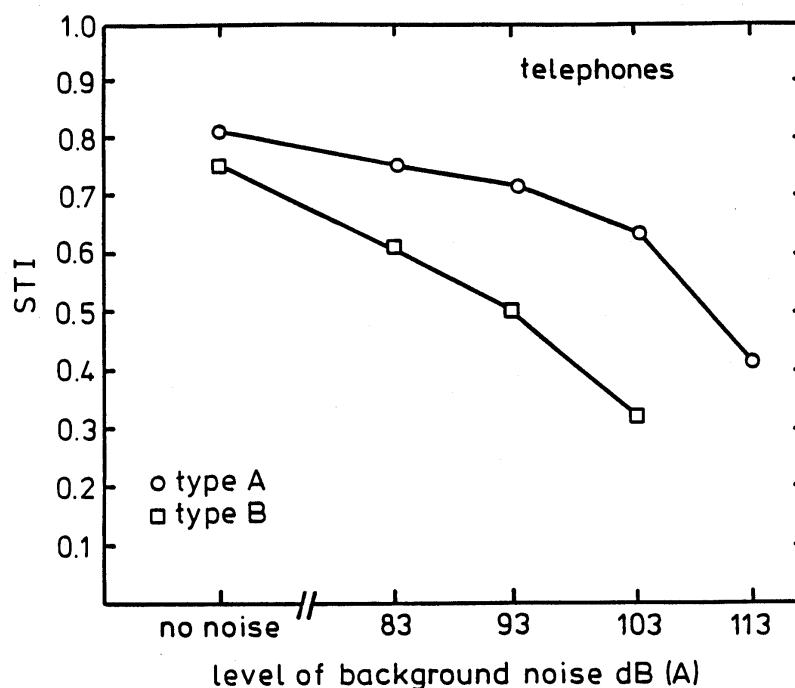


Fig. 8.6 STI, for two types of headset as a function of the background noise level. The presentation level of the test signal was 75 dBA.

8.3 Room acoustics and public address systems

Measurements in auditoria or with public address systems are normally performed with the STI-14 or STIPA method. This includes the measurement of the MTF for 14 modulation frequencies. If a smooth MTF is obtained, one can decide to decrease the resolution by skipping modulation frequencies. For some applications it is not necessary to measure the MTF for all the seven frequency bands. The resolution in the frequency domain may be reduced to two octave bands (with a centre frequency of 500 Hz and 2000 Hz). This is only valid when no limitation in frequency transfer is effective (no PA-systems) and the background noise is of minor importance or can be described by samples within these two frequency bands. The RASTI method is an example of an application with these limitations.

An example of the use of the RASTI method is given in Fig. 8.7. For a number of positions in an auditorium the STI was measured and the results were plotted in a lay-out of the room. Adjacent measuring points with a similar STI-value were connected. This results in iso-STI contours. The contours are usually made at intervals of 0.05 STI. A high gradient of the STI indicates a poor distribution of the intelligibility in the room.

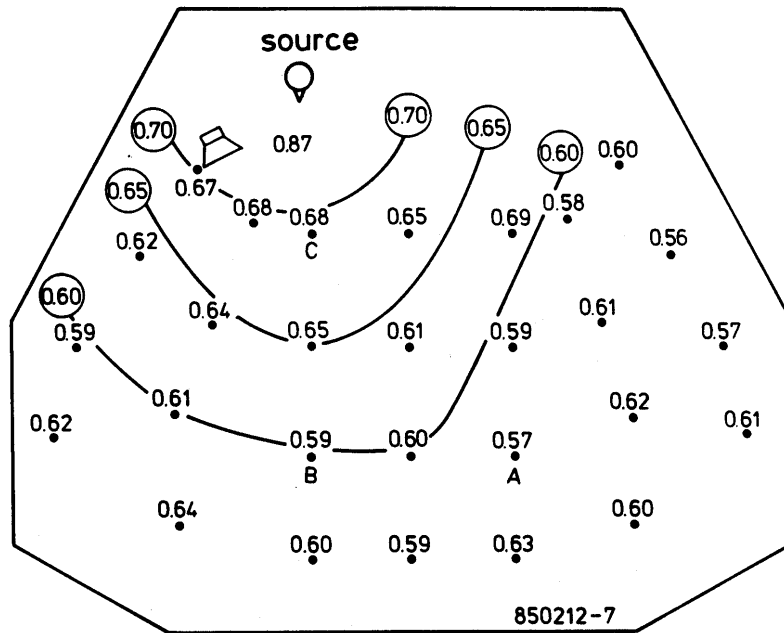


Fig. 8.7 Iso-STI contours for an auditorium with no background noise.

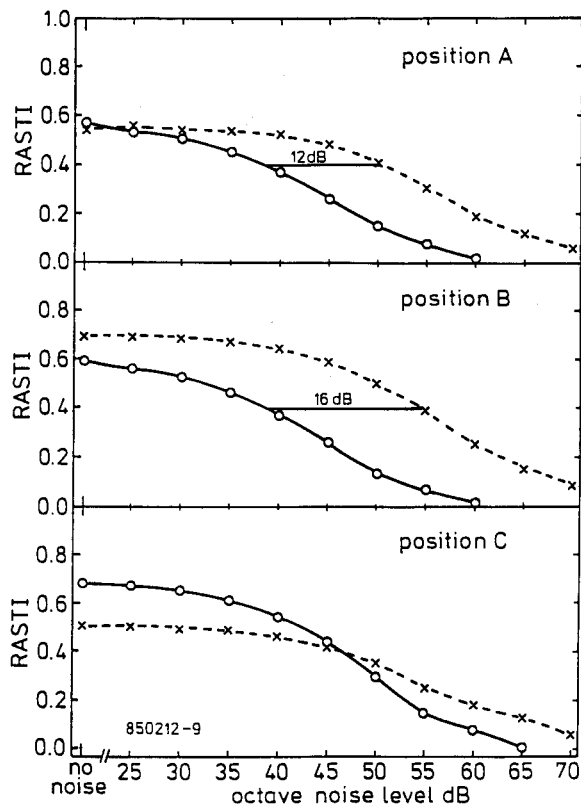


Fig. 8.8 STI as a function of the background noise level for three positions in the auditorium of Fig. 8.7 and with the PA-system switched on and off.

If a PA-system is used, this system may increase the direct speech level, but it increases also the level of reverberating speech sound. This depends on the directivity and positioning of the loudspeaker(s) and the presence of sound absorbing material (e.g., the public). Hence in some cases the use of a PA-system may be beneficial, in other cases it may reduce the speech intelligibility. An example of this effect is given in Fig. 8.8. For three positions in the auditorium of Fig. 8.7 (A, B,C) the full STI is measured as a function of the noise level and for the condition with and without the PA-system. Position B shows an increase and position C shows a decrease of the STI due to the PA-system.

For position A and B the STI as a function of the noise level is increased. The horizontal shift of the two curves (with and without PA-system) shows the *effective gain* (the same STI at higher noise levels). It is obvious that this gain is minimal for position C. This method can be used to optimise PA-systems.

The MTF and the reverberation time in an enclosure are related (theoretically) according to the formula given in section 6 (Fig. 6.2). Hence, based on the measured MTF, the reverberation time T can be estimated. This is demonstrated in Fig. 8.9. In this graph the MTF's measured for several conditions in the same auditorium are given. Two parameters were varied: (a) the use of the PA-system and (b) additional sound absorbing material spread on the floor. The MTF given in the graph is measured within the octave band with centre frequency 2000 Hz. It is shown that for this example the use of the PA-system does not affect the MTF and hence does not change the reverberation time. The use of additional absorbing material however, has a significant effect on the MTF.

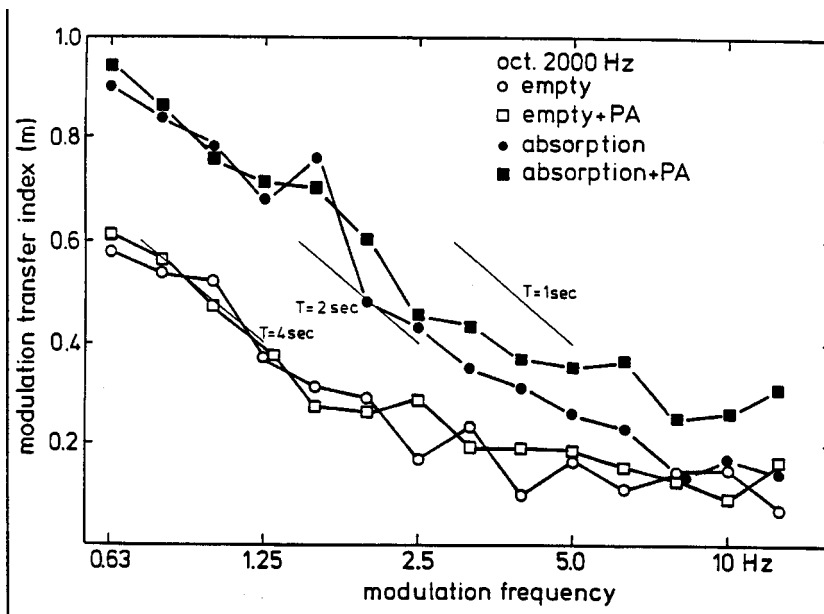


Fig. 8.9 MTF for one position in an enclosure and based on the octave band with centre frequency 2000 Hz. Four conditions are observed being the combination of a PA-system switched on and off and the use of additional sound absorbing material.

In Fig. 8.9 a small part of the theoretical MTF's corresponding to reverberation times of 1, 2, and 4 s respectively are drawn. These curves are calculated according to the formula given in Fig. 6.2. This formula is based on a simple exponential decay curve (no coupled enclosures). The reverberation time is estimated by fitting the measured MTF with the theoretical MTF's. In the example the reverberation time $T = 4$ s for the two conditions without additional absorbing material and approximately $T = 2.2$ s for the condition with the additional absorbing material. It should be noted that the MTF approach is closely related to the perception of fluctuations. The predicted reverberation time is related to the *early decay time* rather than to the conventional reverberation time.

10 REFERENCES

- Anderson, B.W., and Kalb, J.T. (1987). "English verification of the STI method for estimating speech intelligibility of a communications channel," *J. Acoust. Soc. Am.* **81**, 1982-1985.
- ANSI (1969). *Ansi S3.5-1969, American national standard methods for the calculation of the articulation index*, American National Standards Institute, New York.
- Barnett, P. W. and Knight, R.D. (1995). "The Common Intelligibility Scale", *Proc. I.O.A.* Vol 17, part 7.
- Barnett, P. W. (1999). "Overview of speech intelligibility" *Proc. I.O.A* Vol 21 Part 5.
- Berry, R.W. (1971). "Speech volume measurements on telephone circuits," *Proc. IEE* **118**(2), 335-338.
- Bos, C.S.G.M., and Steeneken, H.J.M. (1991). "Phoneme confusions in distorted speech: a diagnostic study," Report IZF 1991 I-4, TNO Institute for Perception, Soesterberg, The Netherlands.
- Brady, P.T. (1965). "A statistical basis for objective measurement of speech levels", *Bell System Tech. J.* **44**, 1453-1486.
- Brady, P.T. (1968). "Equivalent Peak Level: A threshold-independent speech-level measure," *J. Acoust. Soc. Am.* **44**, 695-699.
- Dunn, H.K., and White, S.D. (1940). "Statistical measurements on conversational speech", *J. Acoust. Soc. Am.* **11**, 278-288.
- Egan, J.P. (1944). "Articulation testing methods," OSRD report No. 3802.
- Fairbanks, G. (1958). "Test of phonetic differentiation: The Rhyme Test," *J. Acoust. Soc. Am.* **30**, 596-600.
- Fletcher, H., and Steinberg, J.C. (1929). *Bell Sys Tech. J.* **8**, 806.
- Fletcher, H., and Galt, R.H. (1950). "The perception of speech and its relation to telephony," *J. Acoust. Soc. Am.* **22**, 89-151.
- Fletcher, H. (1953). *Speech and Hearing in Communication* (D. van Nostrand, New York).
- French, N.R., and Steinberg, J.C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90-119.
- Hecker, M.H.L., Bismarck G. von, and Williams, C.E. (1986). "Automatic evaluation of time-varying communications systems," *IEEE Trans. on Audio and Electroacoustics* **AU-16**, 100-106.
- House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D. (1965). "Articulation testing methods: Consonantal differentiation with a closed-response set," *J. Acoust. Soc. Am.* **37**, 158-166.
- Houtgast, T., and Steeneken, H.J.M. (1971). "Evaluation of speech transmission channels by using artificial signals," *Acustica* **25**, 355-367.
- Houtgast, T., and Steeneken, H.J.M. (1973). "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acustica* **28**, 66-73.
- Houtgast, T., Steeneken, H.J.M., and Plomp, R. (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," *Acustica* **46**, 60-72.

- Houtgast, T., and Steeneken, H.J.M. (1984). "A multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria," *Acustica* **54**, 185-199.
- Houtgast, T., and Steeneken, H.J.M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069-1077.
- Houtgast, T., and Verhave, J. (1991). "A physical approach to speech quality assessment: correlation patterns in the speech spectrogram," *Proc. Eurospeech '91, Genova*, 285-288.
- IEC-report (1988). "The objective rating of speech intelligibility in auditoria by the 'RASTI' method," Publication IEC 268-16.
- IEEE (1969). "Speech quality measurements," *IEEE Transactions on Audio and Electroacoustics*, September, 227-246.
- Jacob, K., Steeneken, H.J.M., Verhave, J., and McManus, S., (2001). "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.
- Kryter, K.D. (1960). "Speech band-width compression through spectrum selection," *J. Acoust. Soc. Am.* **32**, 547-556.
- Kryter, K.D. (1962a). "Methods for the calculation and use of the articulation index," *J. Acoust. Soc. Am.* **34**, 1689-1697.
- Kryter, K.D. (1962b). "Validation of the articulation index," *J. Acoust. Soc. Am.* **34**, 1698-1702.
- Kryter, K.D., and Ball, J.H. (1964). "SCIM -- A meter for measuring the performance of speech communication systems," Techn. Doc. report No. ESD-TDR-64-674.
- Kryter, K.D. (1970). *The effects of noise on man* (Academic Press).
- Licklider, J.C.R. (1959). "Three auditory theories," in *Psychology: A Study of Science, Vol. 1*, edited by S. Koch (McGraw-Hill, New York), pp 41-144.
- Licklider, J.C.R., Bisberg, A., and Schwartzlander, H. (1959). "An electronic device to measure the intelligibility of speech," *Proc. Natl. Electronic Conf.* **15**, 329-334.
- Mapp, P. (2001) "Improving the intelligibility of aircraft PA-systems" *Proc Institute of Acoustics, reproduced sounds 17, Stratford-upon-Avon.*
- Miller, G.A., and Nicely, P.E. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338-352.
- Payne, J.A., and McManamon, P.M. (1973). "An objective speech quality measurement of a communication channel," OT report 73-14, Department of Commerce, Office of Telecommunications.
- Pavlovic, C.V., and Studebaker, G.A. (1984). "An evaluation of some assumptions underlying the articulation index," *J. Acoust. Soc. Am.* **75**, 1606-1612.
- Pavlovic, C.V. (1987). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," *J. Acoust. Soc. Am.* **82**, 413-422.
- Plomp, R., Steeneken, H.J.M., and Houtgast, T. (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. II. Mirror image computer model applied to rectangular rooms," *Acustica* **46**, 73-81.

- Pollack, I. (1948). "Effect of high pass and low pass filtering on the intelligibility of speech in noise," *J. Acoust. Soc. Am.* **20**, 259-266.
- Quackenbush, S.R., Barnwell, T.P., and Clements, M.A. (1988). *Objective Measures of Speech Quality* (Prentice Hall, New Jersey).
- Raaij, J.L. van, and Steeneken, H.J.M. (1991). "Digital simulation of speech transmission channels," Report IZF 1991-A7, TNO Institute for Perception, Soesterberg, The Netherlands.
- Rietschote, H.F. van, Houtgast, T., and Steeneken, H.J.M. (1981). "Predicting speech intelligibility in rooms from the modulation transfer function. IV. A ray-tracing computer model," *Acustica* **49**, 245-252.
- Schroeder, M.R., (1981) "Modulation Transfer functions: Definition and Measurement", *Acustica* Vo. 49, pp.179-182.
- Schwartzlander, H. (1959). "Intelligibility evaluation of voice communications," *Electronics* **29**, 88-91.
- Steeneken, H.J.M., and Houtgast, T. (1973). "Intelligibility in telecommunication derived from physical measurements," *Proc. Symp. Intelligibilité de la Parole, Liège*, 73-80.
- Steeneken, H.J.M., and Houtgast, T. (1978) "Comparison of some methods for measuring speech levels," Report IZF 1978-22, TNO Institute for Perception, Soesterberg, The Netherlands.
- Steeneken, H.J.M., and Houtgast, T. (1979). "Measuring ISO-intelligibility contours in auditoria," *Proc. 3rd Symp of FASE on building Acoustics, Dubrovnik*, 85-88.
- Steeneken, H.J.M., and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **67**, 318-326.
- Steeneken, H.J.M., and Agterhuis, E. (1982). "Description of STIDAS II D, Part 1, General system and program description," Report IZF 1982-29, TNO Institute for Perception, Soesterberg, The Netherlands.
- Steeneken, H.J.M., and Houtgast, T. (1982). "Some applications of the Speech Transmission Index (STI) in auditoria," *Acustica* **51**, 229-234.
- Steeneken, H.J.M., and Houtgast, T. (1983). "The temporal envelope spectrum of speech and its significance in room acoustics," *Proc. 11th International Congress on Acoustics, Paris*, **Vol. 7**, 85-88.
- Steeneken, H.J.M., and Houtgast, T. (1986). "Comparison of some methods for measuring speech levels," Report IZF 1986-20, TNO Institute for Perception, Soesterberg, The Netherlands.
- Steeneken, H.J.M. (1987a). "Diagnostic information of subjective intelligibility tests," *Proc. IEEE ICASSP, Dallas*, 131-134.
- Steeneken, H.J.M. (1987b). "Comparison among three subjective and one objective intelligibility test," Report IZF 1987-8, TNO Institute for Perception, Soesterberg, The Netherlands.
- Steeneken, H.J.M., and Houtgast, T. (1991). "On the mutual dependency of octave-band specific contributions to speech intelligibility," *Proc Eurospeech '91, Genova*, 1133-1136.
- Steeneken, H.J.M. (1992). "Quality evaluation of speech processing systems," Chapter 5 in *Digital Speech Coding: Speech coding, Synthesis and Recognition*, edited by Nejat Ince, (Kluwer Norwell USA), 127-160.
- Steeneken, H.J.M., and Houtgast, T. (1999) "Mutual dependence of the octave-band weights in predicting speech intelligibility". *Speech communication*, 1999, vol.28, 109-123.

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

Steeneken, H.J.M., and Houtgast, T. (2002b). "Validation of the revised STI_r method". Elsevier Speech Communication, 2002, vol. 38.

Studebaker, G.A., Pavlovic, C.V., and Sherbecoe, R.L. (1987). "A frequency-importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130-1138.

Studebaker, G.A., and Sherbecoe, R.L. (1991). "Frequency-importance and transfer functions for recorded CID W-22 word lists," J. Speech Hear. Res., 34, 427-438.

Voiers, W.D. (1977a). "Diagnostic evaluation of speech intelligibility." In *Speech Intelligibility and Speaker Recognition*, Vol. 2. Benchmark papers in Acoustics, edited by M.E. Hawley (Dowden, Hutchinson, and Ross, Stroudsburg), 374-384.

Voiers, W.D. (1977b). "Diagnostic acceptability measure for speech communication systems," Proc. IEEE ICASSP, Hartford CT, 204-207.

Wattel, E., Plomp, R., Rietschote, H.F. van, and Steeneken, H.J.M. (1981). "Predicting speech intelligibility in rooms from the modulation transfer function. III. Mirror image computer model applied to pyramidal rooms," *Acustica* **48**, 320-324.

Wijngaarden, S.J. van, Steeneken, H.J.M. (1999) *Objective prediction of speech intelligibility at high ambient noise levels using the Speech Transmission Index* Proc Eurospeech99, Budapest, 2639-2642.

Zwicker, E, and Feltkeller, (1967). *Das Ohr als Nachrichtenempfänger*, (Hirzel Verlag, Stuttgart), 187-200.