Proceedings of the Institute of Acoustics

SUBJECTIVE AND OBJECTIVE SPEECH INTELLIGIBILITY MEASURES

Herman J.M. Steeneken, Tammo Houtgast

TNO-Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

1. INTRODUCTION

Assessment methods for speech communication systems can be divided into three groups: (1) subjective intelligibility measures focused on phonemes, words or sentences, (2) subjective quality measures related to a global impression, and (3) objective measures based on physical aspects of the speech signal or the speech transmission path.

ad 1. Several methods will be discussed for the subjective evaluation of speech transmission systems, especially concerning the scoring method (open or closed response, scaling), the type of speech material (short nonsense words, rhyme words, phonemes or sentences), and the experimental design.

ad 2. Quality rating is a more global method initially developed for assessment of systems performing at a fair-to-high quality level. The relevance of this type of test is discussed in relation to intelligibility measures.

ad 3. Objective methods, in which the transmission quality is derived from physical parameters, offer additional to the prediction of intelligibility also useful diagnostic information.

2. SUBJECTIVE INTELLIGIBILITY MEASURES AND QUALITY RATING

A number of subjective intelligibility tests have been developed for the evaluation of speech communication channels (Fletcher and Steinberg [4]; Egan [2]; Miller and Nicely [12], Fairbanks [3]). In general, the choice of the test is related to the purpose of the study: are systems to be compared or rank-ordered, are systems to be evaluated for a *specific application*, must the *development* of a system be supported, or is the speech *perception* subject of the study? For each type of application a different test may be appropriate. An overview focused on the assessment of speech processing systems is given by Steeneken [22].

Subjective intelligibility tests can be largely categorised by the speech items tested and by the response procedure used. The smallest items tested are at the segmental level, i.e phonemes. Other test items are CV, VC, and CVC combinations (C=consonant, V=vowel), nonsense words, meaningful words, and sentences.

Besides intelligibility scores, speech quality can also be determined by questionnaires or scaling methods, using one or more subjective scales such as: overall impression, naturalness, noisiness, clarity, etc. Speech quality assessment is normally used for communications with a high intelligibility, for which most tests based on intelligibility scores cannot be applied because of ceiling effects. The overview given below describes representative tests from this segmental level up to sentence level, as well as tests giving a general impression of transmission or speech quality.

Tests at phoneme and word level

A frequently used test for determining phoneme scores is the rhyme test. A rhyme test is a forced-choice test in which a listener, after each word that is presented, has to select his response from a small group of visually presented alternatives. In general, the alternatives only differ with respect to the phoneme at one particular position in the test word. For example, for the Dutch language and for a test with a plosive in the initial consonant position, the possible alternatives might be: Bam, Dam, Pam, Tam, Kam. A rhyme test is easy to apply and does not require much training of the listeners. Frequently used rhyme tests are the Modified Rhyme Test (MRT, testing consonants and vowels) and the Diagnostic Rhyme Test (DRT, testing initial consonants only).

The MRT is based on six alternatives (Fairbanks [3]), the DRT is based on two alternatives (Voiers [25]; Peckles and Rossi [13]; Sotscheck [15]; Steeneken [17]). For the DRT the alternatives are based on testing single articulatory features mainly according to the concept defined by Miller and Nicely [12]. Studies have shown that the DRT, because of the limited number of alternatives, is less sensitive and may force listeners to respond differently from their perceptual impression (i.e the phoneme actually heard by the listener might not be included in the two alternatives presented by the test, Steeneken [21]; Greenspan [6]).

A more general approach is obtained with a test with an open response, such as with monosyllabic word tests (Fletcher [4]; Egan [2]). Open response tests make use of short nonsense or meaningful words of the CVC type. Sometimes VCV words, CV words, VC words, CCVC words, or CVCC words are used. This may depend on features of the particular language (for example Italian has no closed syllables) or the wish to evaluate specific clusters such as consonant clusters or diphone clusters. With nonsense words and an open response, the listener can respond with any combination of phonemes corresponding to the type of word as defined beforehand. This procedure requires extensive training of the listeners.

The test results can be presented as phoneme scores and word scores but also as confusions between the initial consonants, vowels, and final consonants.

The confusion matrices obtained with open response tests provide useful (diagnostic) information for improving the performance of a system, Steeneken [20]. Multidimensional scaling techniques may help to visualize the relations between the stimuli.

With word tests it is recommended to embed the words in a *carrier phrase*. Such a carrier phrase (which is neglected in many studies) will cause representative echoes and reverberation in conditions with a distortion in the time domain. Also automatic gain control (AGC) settling will be established by the carrier phrase. An important aspect of using a carrier phrase is also that it stabilizes the vocal effort of the speaker during the pronounciation and that it reduces the vocal stress on the test words. Finally it can function as a cue to the listener that the next test word is going to be presented.

Tests at sentence level

Sentence intelligibility is sometimes measured by asking the subjects to *estimate* the percentage of words correctly heard on a 0-100% scale. This scoring method tends to give a wide spread among listeners. Sentence intelligibility saturates to 100% at poor signal-to-noise ratios, the effective range is small (see Fig. 1).

The speech reception threshold (SRT) measures word or sentence intelligibility against a level of masking noise. The listener has to recognize a word or sentence presented at a fixed level and masked by noise at a variable level. After a correct response the noise level is increased, while after a false response the noise level is decreased. This procedure leads to an estimation of the noise level where a 50% correct identification of the words or sentences is obtained (Plomp and Mimpen [14]). The quality of the speech (and/or of the listener) is related to the amount of noise required for masking. The procedure has the advantage that it can be performed with naive listeners and gives very reproducible results. The standard deviation of the masking noise level for repeated tests with the same speaker and listener is close to 1.5 dB.

Recently the relation between the SRT scores and phoneme and word scores was studied. It was found that CV and CVC words are good predictors for sentence intelligibility but scores based on consonants or vowels only give a poor prediction [23].

The use of anomalous sentences gets a great deal of attention in combination with the assessment of speech synthesis systems. These syntactically correct but semantically anomalous sentences consist of approximately seven words. The words are taken from sets of common monosyllabic words from which a virtually unlimited number of sentences can be generated randomly according to some predefined grammatical structures. The robustness of this test has not yet been shown, but it is being studied in a current assessment program.

Quality rating

Quality rating is a more general method, used to evaluate the user's acceptance of a transmission channel or speech output system. The claim of some investigators, Goodman and Nash [5], is that a quality rating reflects the total auditory impression of speech by a listener and can be used to discriminate between a number of intervals ranging from excellent to bad. For quality ratings, normal test sentences or a free conversation are used to obtain the listener's impression. The listener is asked to rate his impression on a subjective scale such as the five-point scale: bad, poor, fair, good, and excellent. Different types of scales are used, including: intelligibility, quality, acceptability, naturalness etc. Quality rating or the so-called Mean Opinion Score (MOS) gives a wide variation among listener scores (Steeneken [21]). The MOS does not give an absolute measure since the scales used by the listeners are not calibrated. Therefore the MOS can be used only for rank-ordering conditions. For a more absolute evaluation, the use of reference conditions is required as an anchor.

Other methods (more or less related to the scaling methods used with the MOS) are paired comparison, categorial and magnitude estimations.

Relation between various measures

Fig. 1 gives, for five intelligibility measures, the score as a function of the signal-to-noise ratio of speech masked by noise (Steeneken [21]). This gives an impression of the effective range of each test. The given relation between intelligibility scores and the signal-to-noise ratio is valid only for noise with a frequency spectrum similar to the long-term speech spectrum, which makes the signal-to-noise ratio the same for each frequency band. This is for instance the case with voice-babble. A signal-to-noise ratio of 0 dB then means that speech and noise have an equal spectral density.

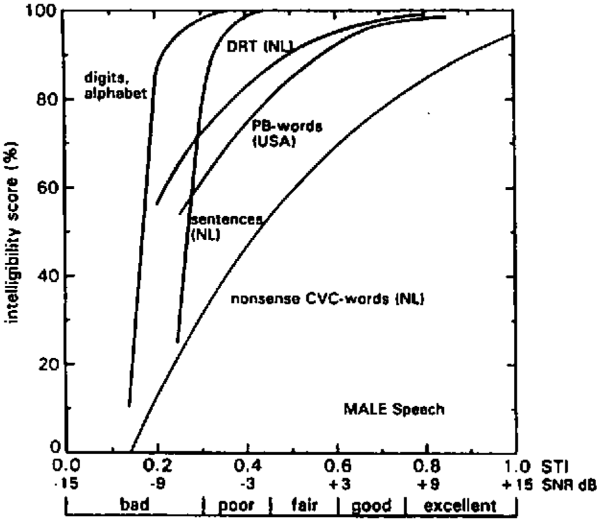

Fig. 1 Qualification of some intelligibility measures and their relation with signal-to-noise ratio for noise with a spectrum shaped according to the long-term speech spectrum and for male speech. Also the relation with the STI (see chapter 3) is given.

As can be seen from the figure, the CVC-nonsense words discriminate over a wide range, while meaningful test words have a slightly smaller range (Anderson and Kalb [1]). The digits and the alphabet give a saturation at a signal-to-noise ratio of -5 dB. This is due to: (a) the limited number of test words and (b) the fact that recognition of these words is controlled mainly by the vowels rather than by the consonants. Vowels have an average level approximately 5 dB above the average level of consonants, and are therefore more resistant to noise. On the other hand nonlinear distortions, such as clipping, will have a greater impact on vowels than on consonants. Therefore the use of the digits and the alphabet, for which recognition is based mainly on vowels, may lead to misleading results. This is indicated in Fig. 2. In this figure the initial-consonant score is given versus the vowel score as obtained from CVC-word tests for 78 different transmission conditions. The graph shows that a high vowel score and a low consonant score can be obtained for one type of channel (e.g. band-pass limiting) while conversely a low vowel score combined with a high consonant score can be obtained for another type of channel (e.g. peak clipping). This indicates that the exclusive use of either consonants or vowels in a subjective test may lead to an incorrect evaluation of the transmission quality. A combination of consonants and vowels, as with CV or CVC words, is required.

The reproducibility of a test strongly depends on the number of speakers and listeners used for the experiments. In general the variance due to the speaker variation is equal to the variance of the listener variations, therefore the number of speakers and listeners should be equal in a test. For CVC based experiments normally four male and four female speakers and 4 listeners are used. This results in 16 speaker-listener pairs for each gender.

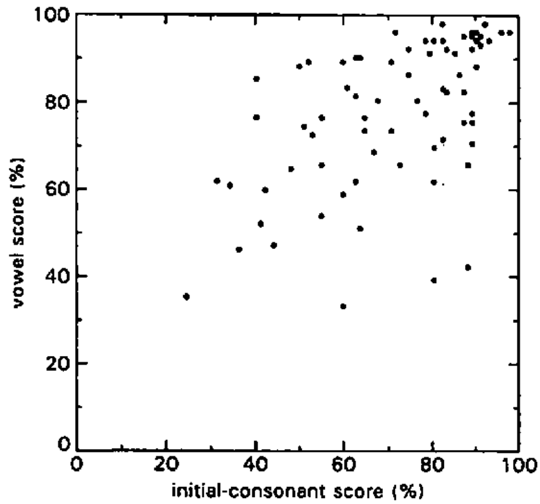

Fig. 2 Initial-consonant score versus vowel score obtained from CVC words for 78 transmission conditions with various combinations of bandwidth, noise, and signal-to-noise ratio.

3. OBJECTIVE INTELLIGIBILITY ASSESSMENT

3.1 Envelope function and envelope spectrum

Connected discourse can be considered as a sequence of the smallest speech items, the phonemes. Each phoneme is represented by a specific frequency spectrum. For the recognition of a speech token the differences between the spectra of the phonemes must be preserved to some degree. These spectral differences are related to the fluctuations in the envelope functions within a number of frequency bands. Distortion of the speech signal, such as noise or reverberation, will result in a reduction of the spectral differences between the spectra of the corresponding phonemes. This is also reflected in the envelope function by a reduction of the fluctuations. In Fig. 3 (panel A) the envelope function, for the octave frequency band 250 Hz, is given.

The shape of the envelope function is unique for a specific sequence of phonemes. A more general description of the fluctuations in the envelope function is given by the *envelope spectrum.* The envelope spectrum results from a 1/3-octave-band analysis of the envelope function (typically of a one-minute speech fragment), and reflects the spectral distribution of the envelope fluctuations relative to the mean intensity: the modulation index as a function of modulation frequency (Fig. 3 panel B). The difference between the original and the resulting envelope spectrum reflects the reduction in the envelope fluctuations caused by the transmission path. This leads to the Modulation Transfer Function (MTF) which represents the reduction factor of the modulation index as a function of modulation frequency.

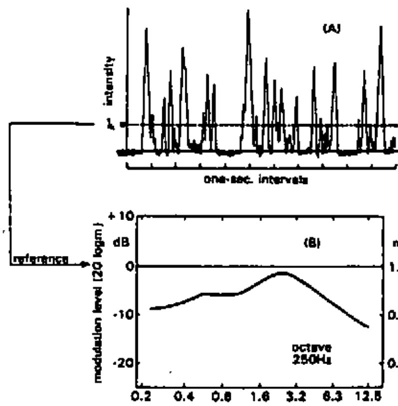

Fig. 3 Envelope function (panel A) of a 10 s speech signal for the octave band with centre frequency 250 Hz. The corresponding envelope spectrum (panel B) is normalized with respect to the mean intensity (I_k).

3.2 The Modulation Transfer Function (MTF)

The rationale underlying the application of the MTF concept in room acoustics has been described in various papers [9,16]. The MTF quantifies to what extent the modulations in the original signal are reduced, as a function of the modulation frequency. The modulations are defined by the *intensity envelope* of the signal: it is in the intensity domain, that interfering noise or reverberation will affect only the degree of modulations of a sine-wave shaped modulation *without* affecting the sine-wave shape. The scheme in Fig. 4 illustrates how the MTF may be used to quantify the relation between the original speech signal at the input and the output signal (A or B). Since most disturbances may vary considerably as a function of frequency, the analysis is octave-band specific. The example of Fig. 4 considers one octave band only, i.e., the intensity envelopes in the octave band with centre frequency of 250 Hz. Two simple sound transmission systems are illustrated, one with reverberation only (case A; T = 2.5 s) and one with interfering noise (case B; signal-to-noise ratio S/N = 0 dB.

Typically, as may be observed in Fig. 4, in the case of reverberation the MTF has the shape of a low-pass filter: the fast fluctuations are relatively more affected by reverberation. In the theoretical case of an ideal exponential reverberation process, the MTF is defined mathematically (see Fig. 3). The typical low-pass character is determined by the product FT (F = modulation frequency, T = reverberation time). In the case of noise interference, the MTF is defined by the *S/N* ratio and is independent of modulation frequency: the interfering noise results in an increased mean intensity and thus reduces the (relative) modulation index for all modulation frequencies by the same factor. In general the shape of the MTF is related to the type of distortion in the time domain. Stationary noise, reverberation and echoes can be identified separately. This gives the MTF some diagnostic properties.

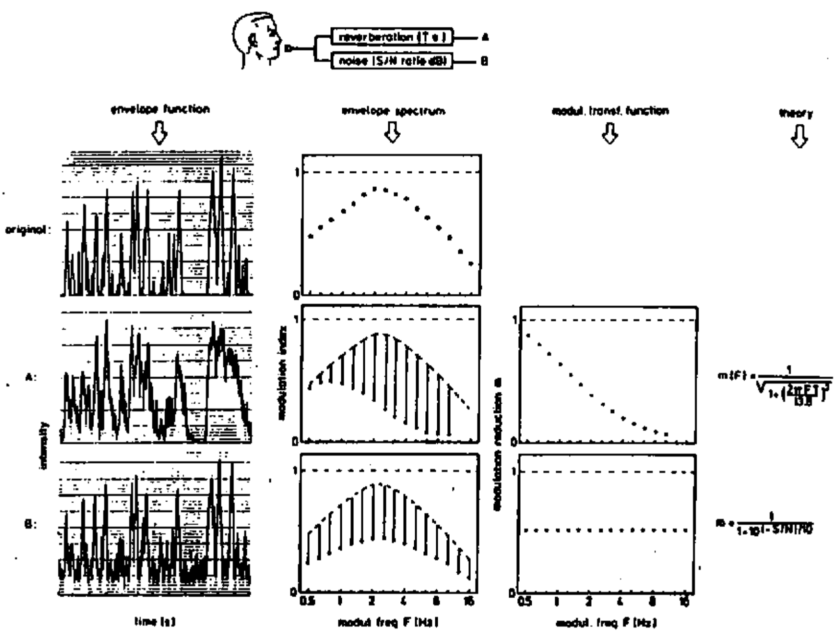

Fig. 4 The reduction of the fluctuations in the (octave-band specific) envelope of an output signal (A or B) relative to the original signal can be expressed by the Modulation Transfer Function. The two conditions considered (reverberation or noise interference), lead to characteristic MTFs, according to the theoretical expressions given at the right-hand side.

It is important to note that the (octave-band specific) MTF of a sound transmission system is *independent* of the input signal considered. It quantifies the modulation transfer for any input signal: speech, music or an artificial signal.

The MTF of a sound transmission system can be determined in various ways, the principle always being that the modulation reduction factor is derived from a comparison of the intensity modulations at the output and at the input of the system. For this purpose either (1) speech signals, (2) the impulse response of a system, or (3) specific test signals can be used.

-1. The use of speech signals has the advantage that the MTF of an enclosure can be determined during a life performance. However, the method is less accurate than the use of artificial test signals [18].

-2. The impulse response can be used to determine the effect of reverberation or echoes on the MTF. *The impulse response method is NOT suited for including the effect of background noise, band-pass limiting and non linear distortion (PA systems) as the average speech spectrum an level distribution is not represented in the test signal.*

-3. The use of an artificial test signal allows the determination of the modulation reduction factor for each modulation frequency successively. In principle, the test signal is produced at the position of the speaker's mouth. It consists of a noise carrier with 100% intensity modulation. The remaining modulation index at a listener's location directly reflects the modulation transfer for that particular modulation frequency. The noise carrier is octave-band filtered, and the measurements are performed for different centre frequencies (typically from 125 Hz up to 8 kHz). For a representative determination of the signal-to-noise ratio, the mean intensity of the test signal should be related to that of the speech as normally produced by a speaker at that position. As a rule, for each octave band considered, the L_{eq} of the test signal is to be adjusted to the L_{eq} of ongoing speech typical for the condition being tested. An example of this measuring scheme is given in Fig. 5.

Based on an adequate test signal, the performance of a sound transmission system can be quantified by a family of MTF curves (ranging from 0.63-12.5 Hz in 1/3 octave steps), comprising $7 \times 14 = 98$ m-values. The question remains of how to transform such a set of data into one single index representing the effect of that transmission system on speech intelligibility: the Speech Transmission Index (STI). The criterion for the relevance of such a transformation is that for a wide variety of transmission systems with different types of disturbances, the relationship between the STI-values and the effect on speech intelligibility is unique, i.e., not system specific. The STI-method based on the full matrix is optimized for (combinations of) distortions as noise, band-pass limiting, reverberation, echoes, and non-linear distortion.

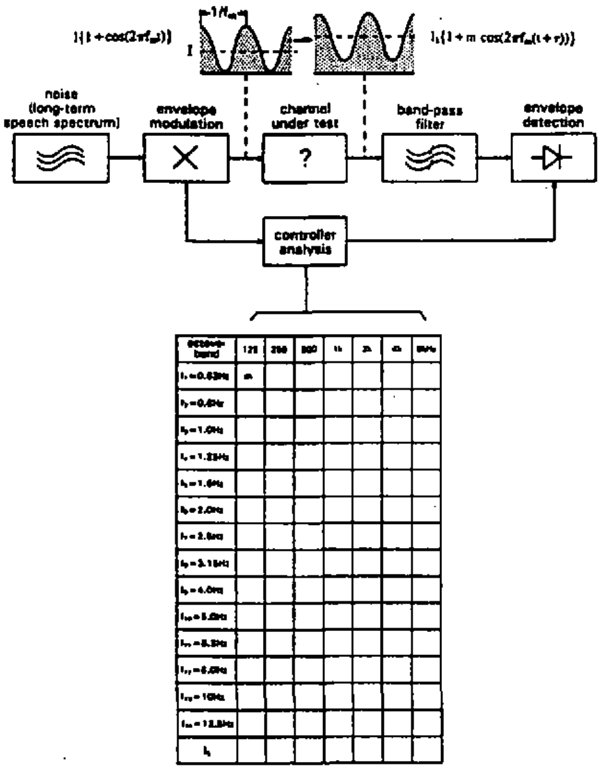

Fig. 5 General block diagram of the measuring set-up. The modulation index reduction at the output (m_o/m_i) is determined for all cells of the matrix (7 octave bands and 14 modulation frequencies). Also the octave levels are obtained, for calculation of the auditory spread of masking.

3.3 The Speech Transmission Index (STI)

The algorithm for conversion of a set of m-values into a STI-value, and the experimental verification on the basis of numerous intelligibility tests, is fully described elsewhere [9,16,24]. An essential step in this transformation is a conversion of each of the 98 *m* values into an *apparent* signal-to-noise ratio $(S/N)_{app}$, irrespective of the actual type of disturbance causing the *m* value. It is interpreted as if it had been caused by interfering noise exclusively, $(S/N)_{app}$ being the signal-to-noise ratio which should have resulted in that *m* value. The conversion is defined mathematically by

$$(S/N)_{app} = 10 \log \frac{m}{1-m} \text{ dB} \qquad (1)$$

being the inverse of the expression given in Fig. 1. A weighted average of the 98 apparent signal-to-noise ratios thus obtained results in the STI, after applying appropriate normalisation such that the signal-to-noise ratios are limited to maximum 15 dB and minimum -15 dB.
By this calculation scheme each family of MTF curves can be transformed unambiguously into a STI value, by which the performance of that sound transmission system is quantified. Also, given the theoretical relations between *m(F)* and the reverberation time *T* or the *S/N* ratio, the calculation scheme may be used for theoretical studies on the effect of reverberation and ambient noise in general.

It has been shown that the STI calculation scheme can be used to predict the performance of an auditorium in the design stage, especially when modelling the sound field along the lines of geometrical acoustics, i.e., by ray-tracing .

There exists a large body of experimental data on the relation between the STI and intelligibility scores obtained with speaker-listener panels [8,16,23]. Typical relations are given in Fig. 1. These relations are only illustrative since, besides the performance of the transmission system, intelligibility scores are affected by other factors also, such as the degree of training and skill of the speaker-listener panel and specific aspects of the speech material employed in the test (e.g., the use of a carrier phrase).

The qualification intervals (bad...excellent) specified along the abscissa in Fig. 1 are based on a large-scale study [8,16], involving various intelligibility tests and different languages.
In the middle range, each qualification interval corresponds to an interval of 0.15 along the STI scale. This implies that differences of that magnitude are important: for two conditions with a STI difference of 0.15, the difference in speech intelligibility is significant and clearly noticeable. Accordingly, for an actual STI-measuring device it is required that the accuracy interval (e.g., the standard deviation for repeated measurements) is considerably smaller than 0.15. This may serve as a guideline for the implementation of the MTF concept in room acoustics along the lines presented in this contribution.

A screening device according to the concept described above has been developed for application in room acoustics. This device, RASTI (RApid-STI) is described in IEC recommendation R268-16 [10]. The RASTI method is suitable for wide-band systems with stationary back-ground noise. Hence, fluctuating back-ground noises, band-pass limitations, and non linear distortions are not accounted for.

For applications on telecommunication channels (and with some limitations also room acoustics) a fast method (measuring time typically 15 s) has been developed. This method is software based and makes use of pre-recorded test signals (CD-ROM) and PC based analysis software [24]. This so-called STITEL-method (STI for TELe communication) is designed for systems with band-pass limiting, noise, non-linear distortion and within certain limits reverberation. Applications are communications systems, public-address systems, electro acoustics, wave-form coders, etc.

4. APPLICATION EXAMPLES

Iso-STI Contours
The normal procedure for determining iso-STI contours is to measure STI values at a large number of positions evenly distributed throughout the audience area. In this way the STI can be mapped and used to construct iso-STI contours.

Depending on the gradient between successive STI values and on the resolution of the measuring grid, iso-STI contours can be drawn for 0.05, 0.1, or 0.2 STI intervals. In Fig. 6 iso-STI contours, based on 29 measuring positions, are given for a lecture hall. In this example, for the empty hall and no background noise, the STI varies from 0.69 to 0.51 which implies an intelligibility rating between good and fair. Normally, the acoustics consultant starts with a measuring session in the absence of an audience, which may result in a non-representative absorption and noise level. The absence of a representative background noise can be compensated for by the application of an artificial noise source during the measurements or by correcting the STI for an imaginary background noise. In the latter case we have to correct the MTFs for a certain noise level.
Fig. 7 shows iso-STI contours obtained from the same measuring data as given in Fig. 6, but corrected for an imaginary background noise with a level of 40 dB in each octave band.

Under the condition presented in Fig. 6, with only small differences of the STI, it is not very relevant to detect areas in the audience with a poor intelligibility. However, for an imaginary background-noise the iso-STI contours change dramatically. As given in Fig. 1 the STI values range from 0.54-0.36 which means that areas with a poor intelligibility can be detected, caused by a low level of the direct sound far from the speaker or far from any reflecting surface.

For three positions in the auditorium (marked A, B, C in Figs. 6 and 7) the STI as a function of the background-noise level is given in Fig. 8 A, B, C (solid lines). These graphs indicate a low signal level at the positions A and B. Besides acoustical measures, such as a reflecting surface behind the speaker, a public-address system (PA) can be applied to increase the level of the (direct) sound.

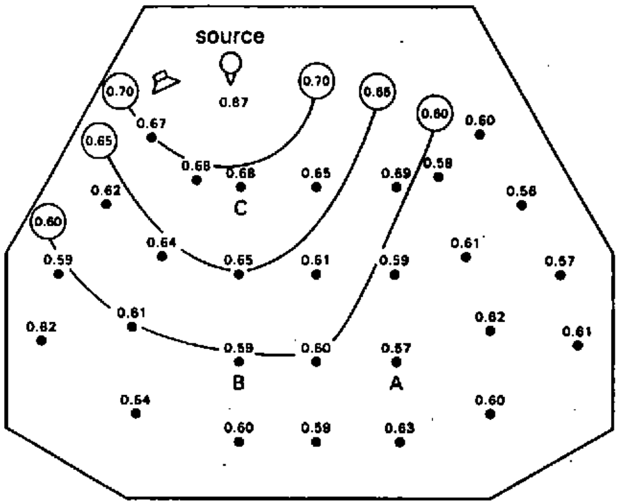

Fig. 6 iso-STI contours for an auditorium, without public, background noise and PA system. The original (29) data points where the contours estimated from, are given as well.

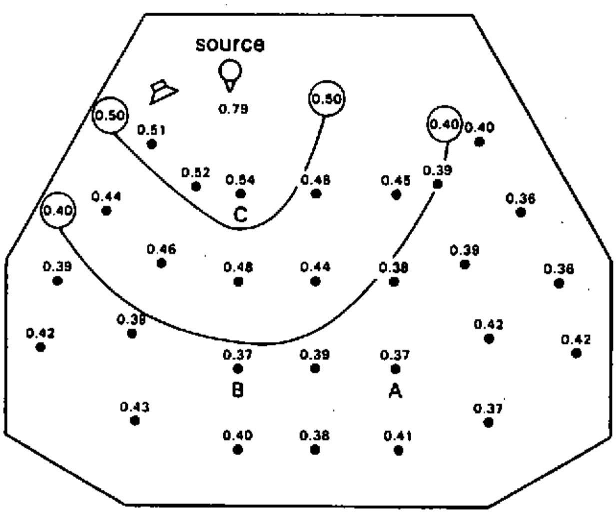

Fig. 7 iso-STI contours for the same data points as given in Fig. 7, but corrected for an imaginary background noise with an octave-band level of 40 dB.

Evaluation of a Public-Address System with the STI

The application of a PA-system in an auditorium increases the direct-sound level at the listener's position and hence the signal's resistance against background noise. The signal level at the listener's position is defined by the system gain and by the position and directivity of the microphone and loudspeakers, and also by the acoustics of the room.

For a poorly designed system, however, with the loudspeakers not optimally directed to the (absorbing) audience, the reverberation field increases, which may result in a decrease of the intelligibility at low noise levels. An example of such a situation is given in Fig. 8 C (dashed curve). For this condition a PA-system in the auditotium as given in Fig. 6 was applied. Only one loudspeaker at the marked position was used. The loudspeaker was placed above the audience, directed to position B.

The STI was measured at positions A, B, and C, and the results, as a function of an imaginary background-noise level, are given in Fig. 8 (dotted lines). For position B the STI value is increased even for the conditions without background noise, which implies a better ratio between the direct and the indirect sound.

We can estimate the contribution of the PA-system by the increase of the resistance against background noise for a given, critical STI-value. As shown in Fig. 8 B this "effective" gain is 16 dB for a STI value of 0.4. For position A this effective gain (of the same PA-system) is 11 dB and at position C it is 0 dB. With this method of validation, an optimal adjustment of the loudspeaker positioning and direction can be found.
In order to exclude the contribution of the microphone and the re-transmission of the indirect sound by the system, the STI-test signal can be connected electrically to the PA-system. As the PA-system may be band-pass limited, the measuring method must be based on the full STI-method or STITEL.

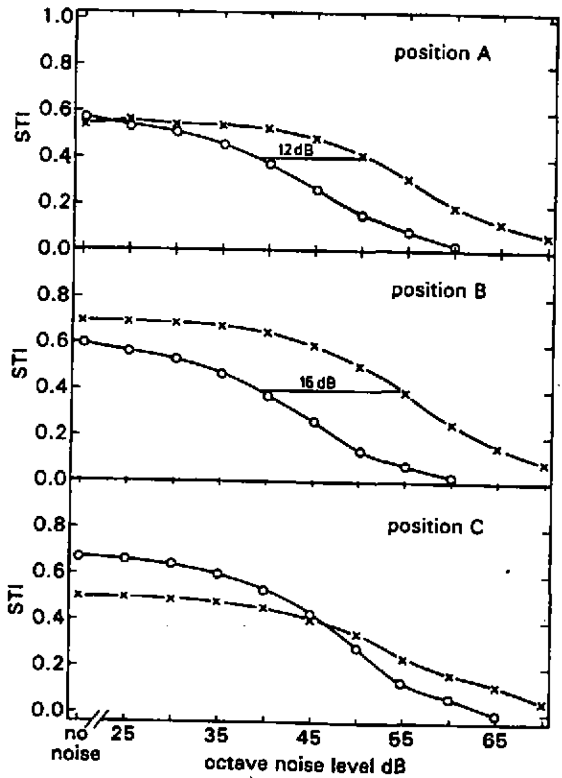

Fig. 8 STI value as a function of an imaginary background-noise level for three positions (marked A, B, C in Fig. 6) and without (o) and with (x) a public-address system. The position of the loudspeaker is also marked in Fig. 6.

Microphone performance in a noise environment

Gradient microphones are developed for use in a high noise environment. The specifications, given by the manufacturers, normally describe the effect of the noise reduction in general terms and are not related to intelligibility, microphone position or type of background noise. In Fig. 9 the transmission quality, expressed by the STI, for two types of microphones is given as a function of the environmental noise level. For these measurements an artificial head was used to obtain the test signal acoustically. The microphone was placed on this artificial head at a representative distance from the mouth. The test signal level was adjusted according to the nominal speech level. This signal level can be increased and the spectrum can be tilted in order to simulate the increase of the vocal effort of a talker in noise (Lombard effect). The head was placed in a diffuse sound field with an adjustable level.

From the figure we can see that the distance from the mouth is an important parameter. It is also obvious that the two noise-cancelling microphones have a different performance in combination with the noise as used in this experiment.

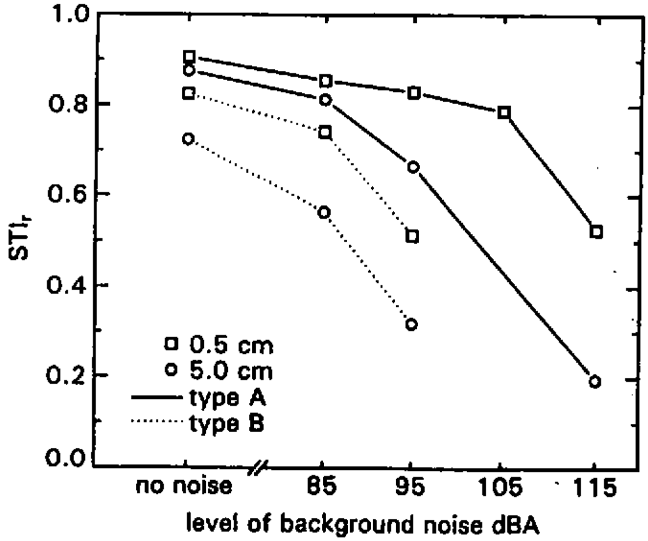

Fig. 9 STI as a function of the noise level for two different microphones and two speaking distances.

5. RÉSUME

Both subjective and objective intelligibility measures were discussed and examples were given. In general the subjective methods, used for room acoustics and public address systems, require much effort. Care should be taken to adapt the test words for adequate use in reverberant environments. Objective measures, i.e. the STI-method, require a simple straight foreward measurement and offer diagnostic information concerning the type of degradation introduced by the transmission path.

6. REFERENCES

[1] 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 (6), 1982-1985.

[2] Egan, J.P. (1944). "Articulation testing methods," OSRD report No. 3802.

[3] Fairbanks, G. (1958). "Test of phonetic differentiation: The Rhyme Test," J. Acoust. Soc. Am. 30, 596-600.

[4] Fletcher, H., and Steinberg, J.C. (1929). Bell Sys Tech. J. 8, 806.

[5] Goodman, D.J. and Nash, R.D., (1984). Subjective quality of the same speech transmission conditions in seven different countries, IEEE Trans Comm. 30, 642-654.

[6] Greenspan, S.L., Bennett, R.W. and Syrdal, A.K., (1989). A study of Two Standard Speech Intelligibility Measures. Presented 117th Meeting Acoust. Soc. Am.

[7] 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.

[8] Houtgast, T. & Steeneken, H.J.M., (1984). "A Multi-Language Evaluation of the RASTI-Method for Estimating Speech Intelligibility in Auditoria", *Acustica* 54 (1984), 185-199.1

[9] Houtgast, T. & 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. Amer.* 77, 1060-1077.

[10] IEC-report.(1988). The objective rating of speech intelligibility in auditoria by the "RASTI" method, Publication IEC 268-16.

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

[12] Miller, G.A., and Nicely, P.E. (1955). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. 27, 338-352.

[13] Peckels, J.P. and Rossi, M., (1973). Le test diagnostic par paires minimales. Revue d'Acoustique No 27, 245-262.

[14] Plomp, R. and Mimpen, A.M., (1979). Improving the reliability of testing the speech reception threshold for sentences, Audiology 8, 43-52.

[15] Sotscheck, J. (1982). "Ein Reimtest fur Verständlichkeitsmessungen mit Deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachguteübertragung," Fernmeldeingenieur 36, 1-84.

[16] Steeneken, H.J.M. and Houtgast, T., (1980). A physical method for measuring speech-transmission quality. J. Acoust. Soc. Am. 67 (1), 318-326.

[17] Steeneken, H.J.M., (1982). Ontwikkeling en toetsing van een Nederlandstalige diagnostische rijmtest voor het testen van spraakkom-munikatiekanalen. Report IZF 1982-13, TNO Institute for Perception, Soesterberg, The Netherlands.

[18] Steeneken, H.J.M. & Houtgast, T., (1983). "The Temporal Envelope Spectrum of Speech and its Significance in Room Acoustics", in: *Proceedings 11th International Congress on Acoustics,* Paris, Vol. 7, 85-88.

[19] Steeneken, H.J.M. & Houtgast, T., (1985) "RASTI: A Tool for Evaluating Auditoria", *B&K Technical Review* 3, 13-30.

[20] Steeneken, H.J.M., (1986). Diagnostic information of subjective intelligibility tests. Internat. IEEE Proc., ICASSP, Dallas.

[21] Steeneken, H.J.M. (1987). "Comparison among three subjective and one objective intelligibility test," Report IZF 1987-8, TNO Institute for Perception, Soesterberg, The Netherlands.

[22] 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.

[23] Steeneken, H.J.M. (1992). "On measuring and predicting speech intelligibility" Doctoral thesis Univ. of Amsterdam.

[24] Steeneken, H.J.M., Verhave, J. and Houtgast, T. (1993) "Objective Assessment of Speech Communication Systems; introduction of a software based procedure". Proc Eurospeech '93 Berlin, 203-206.

[25] Voiers W.D., (1977). Diagnostic Evaluation of Speech Intelligibility. Chapter 32 in M.E. Hawley (ed.) Speech Intelligibility and speaker recognition, Vol. 2. Benchmark papers in Acoustics, Dowden, Hutchinson, and Ross, Stroudburg, Pa.