Past Present and Future of the STI

A explanation of the powerpoint presentation: **Past_present_future_STI-2014.pps**

Slide 3:
In general the reduction of the speech intelligibility is related to a reduction of the signal-to-noise ratio. Hence, speaking louder, increasing the directivity factor of the listener or a well designed public address system might help.

Slide 4:
Assessment of the intelligibility of a speech communication channel can be performed by subjective and objective test methods.

Subjective evaluation: By making use of a speaker, representative speech material has to be transmitted through the channel under test (such as sentences or selected test words). A listener at the receiving side has to write down the test words or just score his/her impression (sentences) of the intelligibility. For a representative test at least 4 speakers and 4 listeners should be used.

Objective assessment: Objective methods determine various physical properties of the channel under test and predict a score related to the intelligibility.

Slide 5:
The graph shows the average speech spectrum and the spectrum of a (white) noise. For 7 octave bands the signal-to-noise ratio (SNR) can be determined. The graph shows a positive SNR for frequency bands up to 2 kHz and a negative SNR for the 4 kHz and 8 kHz band.
Each SNR value is converted to an index (0-1) such that an SNR of -15 dB relates to an index of "0" and an SNR of +15 dB relates to an index of "1".
As each octave band has a different contribution to intelligibility a weighted summation is applied to derive the final STI.

So far the method is valid for speech combined with a stationary noise.

Slide 6:
The envelope function of a speech sample shows how well the speech signal is preserved. Each syllable is represented by a peak in the envelope function. The envelope function is unique for each sequence of words. However, if we determine the frequency spectrum of the envelope function we will get a more general description which is a reproducible measure for longer speech tokens (one minute).
In the lower graph (B) the envelope spectrum is given for the envelope function (A) of the octave band 250 Hz. The spectrum is referred to the average band level.

Slide 7:
Temporal distortions (echoes, reverberation, and automatic gain control) reduce the intelligibility of speech. For example reverberations mask fast fluctuations in the speech signal.
This is similar to a poor spatial resolution in a visual display. The spatial line-frequency (number of lines per cm) varies for the vertical lines in the centre of the graph. If a camera is out of focus (or the band-width of a TV-channel is limited) the lines with the highest spatial frequency will merge and a gray area rather than distinct separate lines will be displayed. This degradation as a function of the spatial line frequency is called the Modulation Transfer Function. A similar approach can be used for temporal distortions in room acoustics (see next slide).

Slide 8:
Here we see the effect of two types of distortion on the envelope function and envelope spectrum. The upper graph shows the original envelope function and envelope spectrum.
The speech signal in graph A was masked by reverberation. Fast fluctuations are smeared, slow fluctuations remain (compared with the original). The envelope function shows a decrease for the fast fluctuations at higher frequencies. If we subtract the envelope spectrum from the original (differences shown by the vertical lines), we get the MTF for this type of reverberation. The reduction is given by the formulae.
The speech signal in graph B was masked by noise, hence the fluctuations in the envelope will not be disturbed but the average band level will be increased (see bottom of envelope function) consequently the envelope spectrum (given with respect to the average band level) will decrease. The difference is independent of the modulation frequency. The reduction is given by the formulae.

Slide 9:
For the measurement of the MTF an artificial test signal is used. Some of the reasons are to increase the measurement accuracy, to reduce the measuring time and to obtain diagnostic information.
The test signal consists of a modulated noise carrier (intensity modulation 100%, m=1) and modulation frequency F. After transmission through a noisy channel the modulation index will be reduced due to the noise. This reduction is a measure for the SNR. By measurement of m as a function of the modulation frequency F, the MTF will be obtained.

Slide 10.
For a full STI measurement the MTF has to be determined for 7 octave bands (125 Hz – 8 kHz) and for 14 modulation frequencies (0.63 Hz – 12.5 Hz).
Under the orange buttons a sample of a test signal for 1, 3, and 10 Hz is given. If you listen to this signal in a reverberating environment you will notice a decrease of the fluctuations for the higher modulation frequency.

Slide 11:
The STI is derived from the data given in the matrix of the former graph.

Slide 12:
The first step is to correct the measured modulation indices “mkF” for masking by adjacent frequency bands and for the reception threshold. The corrected indices are converted to a corresponding SNR value. This SNR value is limited between -15 dB and +15 dB and then converted to a transmission index (TIk,F). The mean TI (average for each octave-band) is calculated for the 14 modulation frequencies (0.63 – 12.5 Hz). This results in seven Modulation Transfer Indices (MTIk). The STI is obtained by calculation of a weighted combination (ak,ßk) of the seven MTI’s.

Slide 13:
The octave weighting function (ak) depends on the type of speech that has to be predicted. For vowels the mid frequency range (500-2000 Hz) is important while for consonants the higher frequencies (2000-4000 Hz) provide a higher contribution to intelligibility.
This additive model is used for AI and SII and has been improved (1992) for the present STI algorithm.

Slide 14:
Fifty percent of the car presented in this graph is masked. Nevertheless recognition of the type of the car (Mercedes) is easy. Obviously gaps in the presentation will not make the recognition impossible. This may be due to the continuity of the shape of the car, hence some redundancy is available to improve the recognition in a masking condition. With speech a similar effect exists for

the low and higher frequency parts of the speech spectrum. A gap in the frequency transfer does not always reduce the intelligibility. Therefore a revised frequency weighting function was introduced in 1992.

Slide 15:
The frequency weighting function (a, solid line) is given together with a redundancy correction (ß, dotted line). The functions are different for male and female speech. Notice that the frequency range is different for female speech as no energy is observed at the 125 Hz octave band.

Slide 16:
This graph gives the relation between the predicted STIr (subscript "r" means redundancy correction which is the standardized procedure) and the CVC-word score. For this assessment 18 different communication channels (combinations of band-pass limiting and four types of noise) were investigated. The small variance around the best fitting curve (s.d. = 4.4% around the third order polynomial) shows the predictive power of STI for this type of distortion. A similar relation is obtained for female speech.

Slide 17:
The listening tests for the assessment were performed with 4 male and 4 female talkers and two listening panels of 4 listeners. Hence for each gender 4 x 8 = 32 talker-listener pairs were obtained. The picture shows a listening panel in action. The listeners type the test words that they have heard. The responses are automatically processed.

Slide 18:
The test words consist of a combination of a consonant-vowel-consonant. Lists of 51 words are used. Each list is based on an equally balanced selection of 17 initial consonants, 15 vowels, and 11 final consonants. The test words are embedded in a carrier phrase. Carrier phrases are used: to get the listener's attention,
to control the vocal effect of the talker, and (if required) to induce temporal distortion during the presentation of the test word (echo, reverberation).
A few examples are given.

Slide 19:
Example of the relation between STI and CVC-word score for temporal distortion (echoes, automatic gain control).

Slide 20:
The STI was assessed for various languages in an International Round Robin test (1984, 8 participants). Also a qualification scale was derived from these experiments and five quality intervals were determined. These are represented now in various international standards, ISO9921, IEC 60268-16.

Slide 21:
Example of the performance of a Public Address system in a wide body aircraft.

Slide 22:
Iso-STI contours for an auditorium with no back ground noise. Notice the loudspeaker next to the speaker. Notice also the three marks, A, B, C that will be discussed in the next slide.

Slide 23:
The STI calculation scheme allows to introduce artificial noise (by correcting the measured m-value with respect to the measured test signal level). Hence for a condition measured without background

noise the effect of noise (with any spectral shape and level, octave resolution) can be accounted for. In this graph the STI is given for three positions (A, B, C), noise levels up to 70 dB, and no public address (solid line). The conditions with the public address system switched on show a higher performance in noise conditions but also (at position C) a decrease of the STI at low noise levels. Obviously the PA-system introduces additional reverberation or an echo at this position.

Slide 24:
Example of an STI measurement with speech as test signal. The envelope spectrum is determined for the original speech signal (close to the talker) and after transmission. Based on the difference between the two spectra the modulation transfer is obtained. A comparison with an MTF measurement is also given (direct).

Slide 25:
The full STI requires 7 octave bands, 14 modulation frequencies, and a specific (speech-like) modulation for octave bands not under test. With this full-STI measurement a wide scope of distortions can be assessed accurately (band-pass limiting, noise, non-linear distortion, and temporal distortion). STI-3 offers a limited resolution in the time domain (temporal), however the measuring time is reduced.

Slide 26:
In the late 70's a specific version (based on the simple microprocessor 6502) was developed for room acoustics (see ref 7). We called it RASTI (Room Acoustical STI). The reduction to two octaves was valid for person-to-person communications. Hence PA-systems, specific (not contiguous) noise spectra are not specified in the application of RASTI. The present state-of-the-art has made it possible to extend the power of a fast screening system such as STI-PA (2001).

Slide 27:
STI-PA has a higher resolution in the frequency domain and covers the full range of the MTF (in six frequency bands). Therefore it is more suitable for room acoustical assessments. However, for non-linear distortion the full STI (14) is advised.

Slide 28:
Recent developments include binaural hearing, more advanced measurements with speech as test signal, the correction for the effect of non-native talkers and listeners and the use of STI for vocoders.

Slide 29:
Classically binaural effects were not included in STI. A rule of thumb was used (3 dB - 0.1 STI) for improvement by directional hearing. A new method has been proposed (use best ear response for a dummy head, and perform cross-correllogram model, see publication 20, sheet 2.

Slide 30:
CVC-word scores (7 subjects) as a function of the binaural STI as well as the monaural STI. Test conditions were selected to be difficult for the standard STI; the conditions include anechoic conditions,(1-14), a cathedral environment (15-21), a classroom (22-32) and a listening room (33-39). The standard "monaural" CVC vs. STI reference curve is also given.

Slide 31:
CVC-word scores (7 subjects) as a function of the binaural STI as well as the monaural STI. Test conditions were selected to be difficult for the standard STI; the conditions include anechoic conditions,(1-14), a cathedral environment (15-21), a classroom (22-32) and a listening room (33-39). The standard "monaural" CVC vs. STI reference curve is also given.

Slide 32:
Using speech as test signal was developed in the past (see 10). Recent developments with improved signal processing technology make it possible to extend the scope of using speech. The graph represents the relation between STI and CVC scores for various vocoders. It is clear that a good relation is obtained, however a correction of STI 0.3 should be subtracted (see reference 21).

Slide 33:
Review of standards is on ongoing process. In the near future the IEC standard will be reviewed and new developments added.