OBJECTIVE PREDICTION OF SPEECH INTELLIGIBILITY AT HIGH AMBIENT NOISE LEVELS USING THE SPEECH TRANSMISSION INDEX

Sander J. van Wijngaarden and Herman J.M. Steeneken
E-Mail: {vanWijngaarden, Steeneken}@tm.tno.nl
TNO Human Factors Research Institute.
P.O. Box 23,
3769 ZG Soesterberg,
The Netherlands.

ABSTRACT

In many cases the intelligibility of speech in noise may be assumed independent of the absolute sound level; the speech-to-noise ratio (SNR) primarily determines intelligibility. However, at high sound levels, speech intelligibility is found to decrease. Subjective Speech Reception Threshold (SRT) measurements were performed at various speech and noise levels, and with various noise spectra. Decreases in intelligibility between noise levels of 75 and 105 dB(A) were found that correspond to 1 to 3 dB difference in SNR, depending on the noise spectrum. This decrease is not predicted by the standardized Speech Transmission Index (STI), which may be calculated from speech and noise spectra. By introducing level-dependent auditory masking in the STI-calculations, a decrease in intelligibility is predicted that corresponds well to the SRT results.

INTRODUCTION

Speech intelligibility of speech in noise may often be assumed to be independent of the absolute sound level. Speech offering a speech-to-noise ratio of, for example, +5 dB is then expected to yield the same intelligibility at any absolute level. However, at relatively high sound levels (>120 dB), Pollack & Pickett [1] and Kryter [2] reported a decrease in speech intelligibility.

A theoretical reason to expect an influence of the absolute sound level on speech intelligibility is that auditory masking at high levels is different from masking at low levels. Upward spread of masking (masking of higher frequency components by lower frequency components) was found to be larger at higher sound levels in many studies (e.g. [3,4,5]). A stronger upward masking effect is likely to give lower speech intelligibility at the same speech-to-noise ratio. Level dependency of masking is found at levels lower than those at which decreased speech intelligibility is usually expected, and might influence intelligibility at sound levels around or even below 100 dB(A). Such levels are commonly found in some situations where speech intelligibility is both important and critical, such as public address systems in traffic tunnels and intercom headsets in noisy vehicles.

Speech intelligibility under such circumstances may be predicted using objective methods that are based on the effective signal-to-noise ratio. One such method is the Speech Transmission Index (STI). This method does not take the absolute noise level into account. Using the standard STI method [6,7], a noticeable discrepancy between intelligibility predictions and the perceived intelligibility in traffic tunnels is observed. The purpose of this research is determine the extent to which speech intelligibility is degraded at high (but realistic) sound levels, and to modify the STI-model to include the effect of absolute sound level.

The STI-model is mathematically simple an straightforward, motivated by knowledge of speech perception and validated by means of robust psycho-acoustic measurements. Any modification of the STI-model should preserve these characteristics. The introduction of level-dependence in the STI-model has been proposed before [8,9]. This modification is computationally relatively complex, introducing many additional parameters, and does not meet the validation standards upheld in construction of the standard STI-model. Therefore, an alternative modification is sought after.

EXPERIMENTAL SETUP

A robust measure for sentence intelligibility under noisy conditions is the speech-to-noise ratio that corresponds to 50% correct response of short redundant sentences. This measure is known as the Speech Reception Threshold or SRT [10].

In the SRT testing procedure, masking noise is added to test sentences in order to obtain the required speech-to-noise ratio. After presentation of each sentence, a subject responds with the sentence as he or she perceives it, and the experimenter compares the response with the actual sentence. After each correct response, the speech level is decreased by 2 dB; after each incorrect response, the level is increased by 2 dB. The first sentence is repeated until it is responded correctly, using 4 dB steps. This is done to quickly converge to the 50% intelligibility threshold. By taking the average speech-to-noise ratio at the ear over the last 10 sentences, the 50% sentence intelligibility threshold (SRT) is found.

The subjects (listeners) were seated in a silent, acoustically insulated room. A set of Sony MDR-CD770 headphones, capable of producing adequately high sound levels with low distortion, were used to present the recorded sentences, monaurally, to the listeners. Using an artificial head, distortion components at speech and noise levels up to 115 dB(A) were found to be sufficiently small.

Filtering was applied to obtain the required frequency transfer, and to compensate for the frequency response of the headphones.

SRT measurements were carried out using speech of two male and two female speakers. In a standard SRT experiment, speech is masked by noise that has the same long-term spectrum as speech by the corresponding speaker. To investigate the influence of the spectral composition of the noise, other noises were also used: white noise (emphasis on high frequencies), noise with the emphasis on low frequencies (6 dB/octave roll-off above 500 Hz) and simulated noise of fans in a specific traffic tunnel. This 'traffic tunnel noise' is dominated by strong frequency components in the 500 Hz octave band. The noise levels were 75, 85, 95 and 105 dB(A). Four subjects (listeners) participated in each condition.

MEASUREMENT RESULTS

In figure 1, results are given for the standard SRT condition at four different noise levels (one talker, four listeners, masking by 'speech noise').

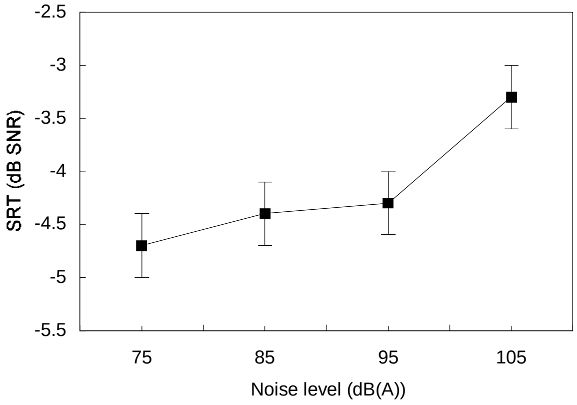

Figure 1. Mean SRT results (SNR corresponding to 50% sentence intelligibility, dB) and standard error (1 talker, 4 listeners) at various noise levels.

At levels of 85 and 95 dB(A), the results are not significantly different from 75 dB(A). At 105 dB(A), there is a small but significant difference.

The subjective difference in speech intelligibility in traffic tunnels, as experienced during informal observations, appears to be larger than explained by the 1.4 dB SNR effect in figure 1. The measurement of figure 1 was therefore repeated with 4 talkers and 4 listeners at 75 and 105 dB(A), but now both with standard noise (long term speech spectrum) and simulated noise with the noise spectrum as measured in a traffic tunnel. Results are given in figure 2.

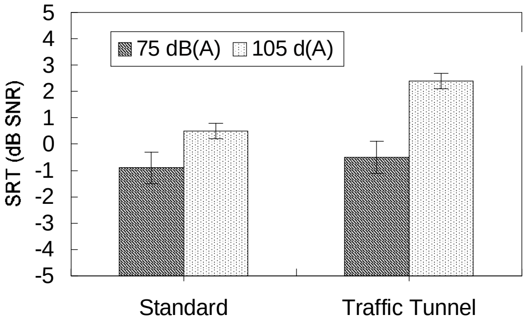

Figure 2 Average SRT results (dB SNR) and standard error (4 talkers, 4 listeners) for two noise conditions.

The overall effect, in terms of difference in SNR between 75 and 105 dB(A), appears to be dependent of the noise spectrum and indeed larger in the traffic tunnel condition. In figure 3, this effect is given for both previous conditions as well as for white noise (+3 dB/octave) and 'low-boost' noise (-6 dB/octave above 500 Hz).

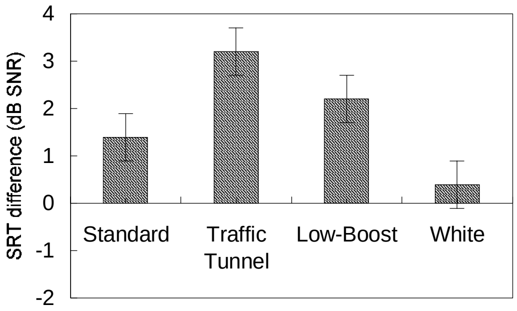

Figure 3. Average difference in SRT results between 75 and 105 dB(A) (4 talkers, 4 listeners) for four noise conditions.

The effect in the white noise condition is very small, and statistically not significant. The fact that the largest effect is found in the traffic tunnel condition (dominated by high levels in the 500 Hz octave band), and the next-largest effect in the low-boost condition, supports the hypothesis that the differences in intelligibility are due to differences in upward masking.

The effects given in figure 3 are large enough to be important for speech intelligibility in practical conditions. It is therefore desirable that objective speech intelligibility measures, such as STI, should predict these effects.

THE STANDARD STI-CALCULATION

Objective speech intelligibility predictions and measurements were obtained using the STI_r-method, as described in the standardization document [7]. The STI_r -method (Redundancy Speech Transmission Index) is an improved version of the classical STI [6], that takes the redundancy between neighbouring frequency bands into account. The improvements incorporated in the STI_r calculation are beyond the scope of this paper, and not relevant to the principles behind auditory masking at high sound levels. We commonly use the general term STI, although all calculations were in fact performed according to the STI_r –method.

The STI-method assumes that the intelligibility of a transmitted speech signal is related to the preservation of the original spectral differences between the speech sounds. These spectral differences may be reduced by bandpass limiting, masking noise, non-linear distortion and distortion in the time domain (echoes, reverberation, automatic gain control). The reduction of these spectral differences can be quantified by the effective signal-to-noise ratio, obtained for a number of relevant frequency bands.

The STI-method is based on calculations of effective signal-to-noise ratios for all relevant frequency bands (seven octave bands, ranging from 125 Hz to 8 kHz). A weighted contribution of the quantified information transfer in the seven octave bands results in a single index, the STI. The measuring method and the algorithm have been optimised for an optimal correlation with subjective intelligibility , which has been validated in a wide range of experiments [11,12].

In this paper, STI predictions and measurements are based on either male or female speakers; SRT results by male speakers are compared to STI results for male speech, and SRT results by female speakers are compared to STI_r predictions for female speech

A STI value of 0.35 is generally presumed to correspond with the Speech Reception Threshold, i.e. the 50% intelligibility level of redundant sentences.

RELATION BETWEEN STANDARD STI AND SRT

At the SNR given by an SRT experiment, speech intelligibility is by definition constant (50% sentence intelligibility). STI measurements at this SNR should therefore always give STI=0.35, irrespective of the condition.

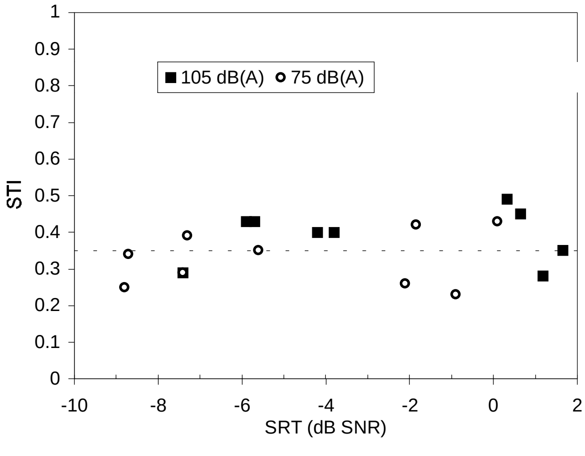

Figure 4. Standard STI measurement results plotted against SRT results, at 75 and 105 dB(A) (2 male or female talkers, 4 listeners) for four noise conditions and a condition with bandwidth-limited speech.

In figure 4, STI results are given for the four noise conditions of figure 3, at the SNR determined by the SRT experiments. Additionally, results are included for SRT and STI measurements with bandwidth limited speech in traffic tunnel noise. Results are separated for male and female talkers.

The spread is considerable, both at 75 and 105 dB(A). Roughly, STI values at 50% sentence intellgibility range between STI=0.25 and STI=0.50, depending on the specific condition. Figure 4 does indicate that the effect of absolute sound level leads to systematically deviating STI-results; the average STI across conditions at 75 dB(A) is STI=0.34, and at 105 dB(A) STI=0.41. STI values at higher sound levels lead to a systematic overestimation of speech intelligibility. This systematic error is of the same order as the statistically estimated error across conditions, observed in figure 4.

MODIFICATION OF THE STI-CALCULATION

The upward spread of masking is accounted for in the STI-model by a single factor: each octave band is presumed to mask the neighbouring higher band with its own level minus 35 decibels. This factor (35 dB) is not level dependent; a logical and simple method to include level dependent masking in the STI-model is by making this factor level-dependent.

Carter & Kryter [3] reported experimental data on the relation between the level of masking noise and the upward spread of masking. The reported masking slope is dependent on both frequency (50 to 5200 Hz) and level (46 to 96 dB SPL); the level dependency (0 to 45 dB/octave) is markedly larger than the frequency dependency (0 to 15 dB/octave).

Table I gives the upward masking slopes by Carter and Kryter as a function of level, for the 50-800 Hz frequency interval. Since upward spread of masking is most important for the lower octave bands, this frequency interval is chosen.

Table I. Upward masking slopes after Carter & Kryter [3] for the 50-800 Hz frequency interval

Level range (octave band level)	Upward masking slope
96-	10 dB/octave
86-95	15 dB/octave
76-85	20 dB/octave
66-75	25 dB/octave
56-65	35 dB/octave
46-55	40 dB/octave

The masking slopes in table I can easily be integrated in the STI-calculation. This adds only few parameters to the STI-model, since table I may be described by a straight line on a certain sound level interval (46-96 dB), and fixed values outside this interval (resp 40 and 10 dB/octave).

The STI-results of figure 4 can be recalculated using the masking curves of table I, for each octave band. The results are given in figure 5.

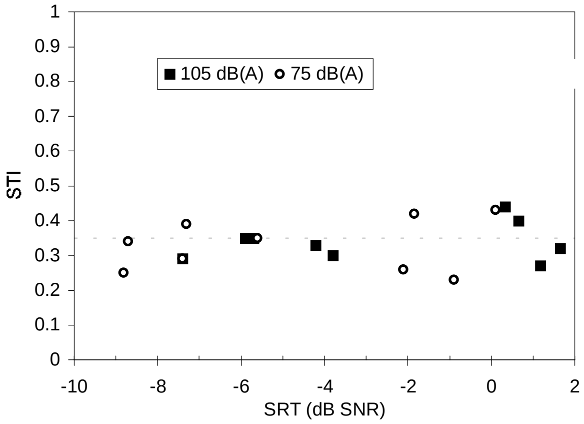

Figure 5. Modified STI results (level dependent masking) plotted against SRT results, at 75 and 105 dB(A) (4 talkers, 4 listeners) for four noise conditions and a condition with bandwidth-limited speech.

In a mean sense, the results for 75 and 105 dB(A) are now approximately equal. In figure 6, mean STI values across conditions and standard errors are given for the standard STI calculation (figure 4) and the modified STI calculation (figure 5).

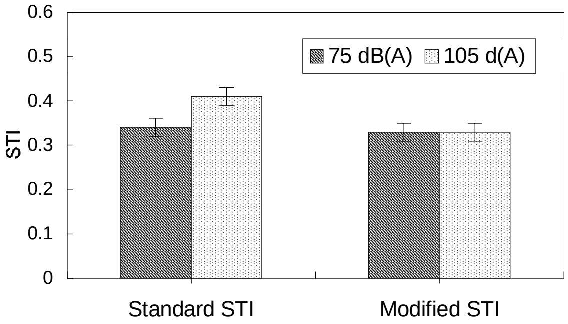

Figure 6. Mean STI results and standard errors for standard STI calculations and the modified STI-model, at 75 and 105 dB(A). Data points taken from figures 4 and 5.

In figure 7, the differences in standard and modified STI values between 75 and 105 dB(A) are given for the data points in figures 4 and 5 corresponding to male talkers.

CONCLUSIONS

At relatively high sound levels (105 dB(A)), a decrease of subjective speech intelligibility was found, equivalent to 3 dB difference in SNR for low-frequency noise. This decrease is relatively small, but large enough to be important in critical situations. The supposed reason for the decrease in speech intelligibility is increased upward masking at higher sound levels.

The Speech Transmission Index (STI) model does not included level-dependency of upward masking. By introducing level-dependent masking slopes from literature, the level-dependency of speech intelligibility is included in the STI. At high sound levels, the modified STI-model is therefore a better predictor of speech intelligibility, at the expense of only a minor, simple modification of the calculation algorithm.

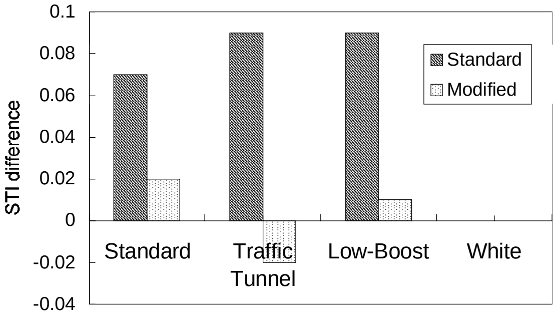

Figure 7. Differences in STI between 75 and 105 dBA (standard and modified model) based on male SRT results of figures 4 and 5 (2 talkers, 4 listeners).

REFERENCES

[1] Pickett, J.M and Pollack, I. (1958). Prediction of speech intelligibility at high noise levels. *Journal of the Acoustical Society of America,* 30,955-963.

[2] Kryter, K.D. (1985). *The effects of noise on man; second edition*. Academic Press, Orlando

[3] Carter, N.L. & Kryter, K.D. (1962). Masking of pure tones and speech. *Journal of Auditory Research,* 2, 66-98.

[4] Zwicker, E. & Feldtkeller, R. (1967) “Das Ohr als Nachrichtenempfänger”, S. Hirzel Verlag, Stuttgart.

[5] Pollack, I. & Pickett, J.M. (1958). Masking of speech by noise at high sound levels. *Journal of the Acoustical Society of America,* 30, 127-130.

[6] Steeneken, H.J.M. & Houtgast, T. (1980). A physical method for measuring speech transmission quality. *Journal of the Acoustical Society of America,* 67, 318-326.

[7] IEC 60268-16 2nd edition (1998) Sound system Equipment ”Part 16: objective rating of speech intelligibility by speech transmission index”, Genève, Suisse.

[8] Buck, K., Wessling, T. & Dancer, A. (1998). Adapting the Speech Transmission Index (STI) for use in very noisy environments. In: Proc. 7th international conference on noise as a public health problem, Sydney, Australia.

[9] Wessling, T. (1997). Erweiterung der Method nach Houtgast und Steeneken zur prognose der Sprachverständlichkeit (sog. STI) für Fälle tieffrequenten Lärms hohen Pegels, Diplomarbeit, Ruhruniversität Bochum.

[10] Plomp, R. & Mimpen, A.M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology,* 18, pp. 43-52.

[11] Steeneken, H.J.M. (1992). On measuring and predicting speech intelligibility. Doctoral dissertation, University of Amsterdam.

[12] Steeneken, H.J.M & Houtgast, T. (1999). Mutual dependence of the octave-band weights in predicting speech intelligibility. *Accepted for publication in Speech Comm.*