Proceedings of the Institute of Acoustics

A COMPARISON OF STI AND ALcons APPLICATIONS IN SPEECH INTELLIGIBILITY ANALYSIS

Paul Doany

Dar Al-Handasah Consultants (Shair & Partners), London

1. INTRODUCTION

Following a brief overview of speech intelligibility analysis in general, this paper concentrates on the Speech Transmission Index *(STI)* and the Articulation Loss of Consonants *(ALcons)* methods. In a first attempt at combining these different methods onto a common set of co-ordinates, the case of pure reverberation is considered where reasonably good agreement.

The effect of the direct field is then considered for the STI method, for which relatively simple expressions are derived enabling computation of the index as a function of reverberation time (T), direct-to-reverberant ratio (D/R) and signal-to-noise ratio (S/N), thereby providing the potential for predictive analysis computations in the traditional way, more in line with the ALcons method. Close agreement is demonstrated with the ALcons measurement algorithm as employed in the Time Energy Frequency Analyser *(Techron/TEF)*, but poor agreement noted with the traditional ALcons predictive formula. The Rapid-STI *(Bruel & Kjaer/RASTI* and the full-STI *(Techron TEF)* measurement algorithms are also briefly considered.

The effect of noise is then assessed, where it is demonstrated that caution should be exercised in interpretation of S/N results from RASTI and TEF.

2. OVERVIEW OF SPEECH INTELLIGIBILITY ANALYSIS TECHNIQUES

2.1 Critical Parameters

Generally recognised causes of reduced intelligibility are: poor S/N; excessive reverberation; specular long-delayed echoes (over 100ms), higher in level than the energy near them [1]. The effects relating to *'masking'* (additive noise) and *'distortion'* (reverberation) are now reasonably well-quantified, however, the effect of *'time-delay'* is much less understood, (although a recent reference by Peutz [2,3] of him being close to an answer is noted).

Other parameters that may be considered in intelligibility analysis include: distance of listener from the source; source directivity factor; aiming with respect to high absorption areas in the listener plane; loudspeaker misalignment between alike devices; misequalisation; non-linear distortion (clipping...).

2.2 Quantitative Methods of Analysis

While there generally appears to be agreement on 'identification' of the main parameters that affect speech intelligibility, controversy still remains on their relative 'quantitative' effect. Smith [4] gives a comprehensive outline of the well-established analysis methods, including: articulation loss of consonants *(ALcons)*, attributed to Peutz; articulation index *(AI)*, attributed to Kryter; modified or equivalent signal-to-noise *(Eq-S/N)*, attributed to Lochner & Burger; modulation transfer function *(MTF)*/speech transmission index *(STI)*, attributed to Houtgast & Steeneken. (Reader is referred to [4] for references to the original paper). It is briefly noted that:

- *ALcons method* gives loss of consonants as a function of T, D/R and S/N; it is restricted to the 2kHz band; 'direct' field computation includes no integration, and hence makes no allowance for "early" reflections, as such; empirically derived modifiers may be added to the original Peutz expression, to account for multiple sources, high absorption in the listener plane, etc.[1]; percentage values do not easily relate to different types of subjective test material, as % ALcons appears to mean different things to different users, although Peutz's comprehensive work in respect of the "information index" should be recognised [2,3];

  AI method gives the articulation index as a function of S/N over the full band; it is suitable for noise-only cases; it provides index values that are very well-established for different types of test material (non-sense syllables, PB-words, sentences...);

  S/N method gives intelligibility as a function of the system impulse response (1000Hz band); it is not considered practical for use in predictive analysis due to complexity of time-weighting but could prove useful in objective measurements; recent work by Bradley [5] and Jacob [6] appears very promising, by illustration of the validity of simple early/late integration by the former, and introduction of the computationally efficient hybrid energy-decay curve (HEDC) by the latter, with the use of computer in the analysis being necessary;

  STI method gives the speech transmission index as a function of modulation transfer function MTF; it is extremely useful in measurements, but is not as easy to use in predictive analysis, not being directly related to the basic parameters of T, D/R and S/N; computer simulation methods using ray-tracing techniques appear very promising especially when combined with statistical methods, as attempted by H.F. Rietschote et Al [7]; the STI method takes the full bandwidth (0.25-8kHz) into account, rather than a limited 1 or 2 kHz band, and offers therefore the potential of giving a complete picture of performance, if it is indeed considered necessary to do so; the serious efforts of Houtgast and Steeneken in relating the index to different types of test material (and indeed languages) must also be recognised [8].

3. MODULATION TRANSFER FUNCTION

Schroeder defines the complex modulation transfer function (CMTF) for linear passive systems [9] as the normalised Fourier transform of the square impulse response (i.e, the energy spectral density function normalised with respect to its value at zero frequency). The real part of this function, the MTF, has been used by Steeneken and Houtgast [9], with it being expressed as a function of modulating frequency F, as follows:

$$m(F) = \frac{\left|\int_0^\infty e^{j2\pi FT} \cdot p^2(t).dt\right|}{\int_0^\infty p^2(t).dt}, \qquad (1)$$

where, p(t) is the instantaneous sound pressure (N/m^2) measured in response to an impulse, in the absence of ambient noise. For example, in the case of pure reverberation (exponential ETC decay), as nominally experienced in the far-field, m(F) may be easily shown [10] to be:

$$m(F) = 1/\sqrt{1 + (2\pi FT)^2}\,. \qquad (2)$$

As only low modulating frequencies F are of interest in the case of speech (0.5-16Hz), it is the slope of the very low frequency response of the magnitude of the energy spectral density function that is of interest in speech intelligibility analysis.

Where the test signal comprises a modulated noise carrier (as in RASTI), thereby enabling also measurement of the S/N effect, the m(F) may be analysed on the basis of the following expression [10]:

$$m(F) = \frac{\left|\int_0^\infty e^{j2\pi FT} \cdot p^2(t).dt\right|}{\int_0^\infty p^2(t).dt} \cdot \frac{1}{1 + 10^{-0.1(S/N)}}. \qquad (3)$$

It is also worth noting here that a slightly modified version of Equation (1) is actually implemented in the TEF analyser as discussed by Keele [12], based upon the Fourier transform of the square magnitude of the complex analytic impulse response (doublet response plus imaginary response). This has no effect on the content of this paper, and is mentioned here for completeness only.

4. SPEECH TRANSMISSION INDEX (STI)

From the MTF, the speech transmission index (STI) may then be computed as defined by Steeneken and Houtgast [10-11], as follows:

(a) Apparent signal-to-noise ratios (S/N)' are computed at n modulating frequencies in each band of interest as appropriate to the measuring device used, and these are then averaged in each band, as follows:

$$(S/N)'\big|_F = 10.\log\left[\frac{m(F)}{1-m(F)}\right], \tag{4}$$

where (S/N)' is hard-limited to +/- 15 dB, and then averaged to yield

$$\overline{(S/N)}' = (1/n)\sum_n (S/N)'\big|_F . \tag{5}$$

(b) STI is computed in each band of interest, as follows:

$$STI = \frac{\overline{(S/N)}' + 15}{30} . \tag{6}$$

An index is computed in each of the seven octave bands from 0.25 to 8kHz inclusive; RASTI however uses the bands 500Hz and 2000Hz, with 4 modulating frequencies in the former and 5 frequencies in the latter, averaged to yield an overall RASTI [11]; i.e., a total of 9 frequencies. The full STI defined by Steeneken and Houtgast, however, requires a total of 98 modulating frequencies over the 7 bands, as realised with the TEF analyser [12].

In arriving at an overall STI, weighting factors are applied to the STI's computed at different octave bands. There appears to be a difference between the weighting factors proposed by Steeneken & Houtgast, and those proposed by Keele [12] as applied in the TEF analyser. Hojberg's implementation using FFT adopts the former weighting factors [13].

5. COMPARISON WITH OTHER METHODS (PURE REVERBERATION CASE)

It would be useful to compare the STI technique with other methods. Figure 1 gives some examples, for the case of pure reverberation. The first set of curves (a,b,c) are plotted against reverberation time (T), including STI, Eq-STI by Peutz [14], and Eq-S/N for 70ms integration, respectively. These would effectively apply in the 1-2kHz band, as necessary.

The second set of curves (x,y,z) are plotted against ALcons (%), including Eq-STI by Becker [15], AI (or S/N) based on ANSI S3.5 [16] (or Harris [17]), and Eq-S/N by Peutz [14], respectively.

The results of these different methods are considered to be fairly close, in the range of 5-25% ALcons, 0.35-.65 STI, and 1 to 4s reverberation time (2kHz). Appendix I gives the equations of the curves discussed.

It is worth noting that the STI/ALcons conversion expression attributed to Becker [15] and followed in this paper coincides almost perfectly with that experienced by Houtgast and Steeneken, for phonetically-balanced nonsense words embedded in a carrier sentence [10].

6. EFFECT OF THE DIRECT FIELD

In the above analysis the effect of the direct field is neglected, assuming operation in the far-field. Houtgast and Steeneken considered the effect of the direct field for a talker [11], for which simple expressions were derived. In this section we derive a more general expression as a function of D/R, essentially based on their work.

If a direct field is to be considered, yet maintaining our earlier exponential decay assumption, $P^2(t)$ may be expressed as follows:

$$p^2(t) = \frac{1}{D^2}\,\delta(t) + \frac{1}{D_c^{\,2}}\,\frac{13.8}{T}\,e^{-13.8(t-t_o)/T}, \qquad (7)$$

where
- D = distance from listener to loudspeaker
- D_c = critical distance = $0.141\sqrt{QR} = 0.057\sqrt{QV/T}$
- T = reverberation time (-60dB)
- Q = directivity factor of loudspeaker (taken along the direction of the listener, assumed on-aim here)
- V = volume
- t_o = initial delay (between direct contribution and reverberant component)
- R = room constant.

Figure 2 gives a representative model of this system, including the ideal reverse Schroeder integration.

Houtgast et Al [11] noted that t_o has little effect on m(F). This depends on D/R and presence of echoes, as we will show in a separate paper. Neverthelesss, neglecting the effect of initial delay t_o:

$$m(F) = \sqrt{\left(1+\beta/(1+\alpha^2)\right)^2 + \left(\alpha\beta/(1+\alpha^2)\right)^2}\;/\;(1+\beta), \qquad (8)$$

where,

$$\alpha = 2\pi FT/13.8 \, , \tag{9}$$

$$\beta = 10^{0.1(D/R)} = (D/D_c)^2, \tag{10}$$

D/R = direct-to-reverberant ratio = 10 log $(D/D_c)^2$

Figure 3a gives the resulting STI as a function of D/R, for various reverberation times, computed for the six modulating frequencies [11]. This applies at any octave band of operation, which could be assumed to be STI 2000, for example.

For the sake of comparison with the ALcons method, results based upon the 2000Hz TEF measurement algorithm [18] are also shown in Figure 3a, which when slightly transcribed to unify symbols may be expressed as follows:

$$\log(\%ALcons) = 2 - 0.64\log(1+10/\beta) + \log(T/12) - 0.32\log(T/12)\log(1+10/\beta) \, . \tag{11}$$

Equation (11) is applicable for S/N above 25dB, and D/R in the range +10 to -12dB; some slight correction is required for values of S/N or D/R outside this range [2]. The Eq-STI value is derived from Equation (11) based upon Becker's formula [15], for plotting in Figure 3a. Figure 3b gives similar curves for readers who prefer to work in percentages. The coherence of these different methods is quite remarkable, especially for D/R values under around +2dB.

Figure 3 gives the ALcons measurement algorithm as applied in TEF [18]. It would also be interesting to include the traditional Peutz prediction formula in this comparison; this is given as curve (a) in Figure 4 for the case of 2s reverberation time [1]. Curves (b) and (c) are identical to those given in Figure 3b, and are included in Figure 4 for comparison, where it is noted that they differ substantially from (a). This has also been confirmed by Barnett and Scarborough [19] under a simulated acoustical environment.

It should be noted that definitions of D/R and T are in accordance with the assumptions made, with no integration for the former, and pure exponential decay for the latter (so that T and EDT are assumed to be the same). Peutz does of-course warn against application of the formula at low EDT (with respect to RT60) [18]. Although the RT60 in TEF should really be EDT [18], based on a 10dB drop in the ETC, some users of TEF have not followed this, and have found it therefore necessary to compensate by allowing some integration in defining D/R. Doany and Mapp [20,21] have experimented with these parameters. The effect of 'early arrivals' in this context is clear, and the debate on how to really quantify such a parameter will undoubtedly continue for some time.

7. EFFECT OF NOISE

This section considers the effect of noise, as the conditions considered in the previous sections assume S/N values above 25dB. The B & K RASTI gives a facility of taking into account the effect of noise on STI. Equation (3) gives the expression for m(F), combining both reverberation (EDT, assumed exponential) and noise. If noise is indeed encountered a RASTI measurement, the effect is a reduction in the modulation index which is flat over the modulating frequency range; in combining the effects of different bands (500/2000Hz), however, the 9dB difference in the transmitted signal level should be duly noted. Alternatively the user may use "noise-floor" entries [22] to derive the expected performance under conditions noisier than those encountered during measurement.

The combined effect of D/R and noise analysis can be easily considered based on Equations (8-10), with the additional multiplier for noise given in Equation (3); some results are plotted in Figure 5, giving ALcons against S/N at D/R values of 0 and -6dB. It is noted that as S/N exceeds 20dB, the curves flatten.

As to the TEF/ALcons algorithm, this is also plotted in the Figure 5 (for the 2kHz band), based on the following expression[2]:

$$\begin{aligned}\log(\%\mathrm{ALcons}) = 2 &- 0.64\log\left(1+10/\beta/(1+\beta')\right)\\ &+ 0.32\log(1+10/\beta').\log(T/12)\\ &- 0.10\log\left(1+10/\beta/(1+\beta')\right).\log(1+10/\beta').\log(T/12)\,,\end{aligned} \qquad (12)$$

where β is as defined earlier, and

$$\beta' = \frac{1+\beta}{\beta}\cdot 10^{-0.1(S/N)} = 10^{-0.1(R/N)}\,. \qquad (13)$$

Equation (12) applies for a limited range of S/N and D/R [18]. It is clear from Figure 5 that there is a shift of around 6dB in S/N between these methods. TEF gives the user the required S/N at which a 10% ALcons expected to be is experienced [18], based on the measured D/R and T; this does not appear to match what RASTI would provide by way of noise floor entries in the same environment.

It is also worth noting that the traditional predictive analysis formula by Peutz for taking S/N into account only applies at the intelligibility distance [1], so this has not been considered here. However, our results do match the work of Metkemeijer et Al (excluding the 2.5% value inc.) [23].

8. SUMMARY AND CONCLUSIONS

In this paper we have attempted to compare STI and ALcons methods under various conditions. It is concluded that very good agreement can be expected under the ideal cases of pure reverberation (T), and presence of a direct field (D/R).

As to noise, caution should be excercised in interpreting results given by TEF and RASTI equipment on S/N, as the results do not appear to match.

Finally, a brief note is included on the expected trends in both predictive analysis and objective measurements, for ALcons, STI and S/N methods. It is considered that the advent of computer modelling and analysis has the potential of improving the latter two methods, with increased interaction between predictions and measurements [24]; the ALcons method however does not appear to benefit as much as its appeal has always been its simplicity. The challenge to the Consultant is to maintain consistency between the various phases of his/her work. Figure 6 illustrates this problem by presenting it in a number of phases: rudimentary acoustical analysis, where simple expressions are manadatory at the inception stage of a project; computer modelling and analysis, the need for which is vital at the advanced design stage; objective measurements and verification testing, where choice of method is crucial if extrapolation of results is required such as in acoustically difficult environments. The appeal of the STI method is lower in the former stage and probably higher in the latter; this paper attempts to initiate a redress of this balance, by highlighting inconsistencies to be pursued in future work.

9. REFERENCES

[1] D. Davis, C. Davis, *"Sound System Engineering"*, pp 235-260, Howard W. Sons & Co., 2nd Edition, USA, (1987).

[2] V. M.A. Peutz, *"Speech Reception and Information"*, App. X, pp 639-644, in above ref [1].

[3] V.M.A. Peutz, "Speech Information and Speech Intelligibility", *AES 85th Convention*, Pre-print 2732 (E-4), Los Angeles, (1988 Nov).

STI. Agreed that answers only validated for Dutch, a Typical language. (Discussion re use of particular modulating frequencies)

[4] H. Smith, "Acoustic Design Considerations for Speech Intelligibility", *J. Audio Eng. Soc.*, Vol. 29, No. 6, pp 408-414, (1981 June).

[5] J.S. Bradley, "Predictors of Speech Intelligibility in Rooms", *J. Acoust. Soc. Am.*, Vol. 80, No. 3, pp 837-845, (1986 Sep).

[6] K.D. Jacob, "Development of a new Algorithm for Predicting Speech Intelligibility of Sound Systems", *AES 83rd Convention*, New York, (1987).

[7] H.F. Van Rietschote, et Al, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function; IV: The A Ray-Tracing Computer Model", *Acustica* 49, pp 245-252 (1981)

[8] T. Houtgast, H.J.M. Steeneken, "A Multi-Lingual Evaluation of the RASTI-Method for Estimating Speech Intelligibility in Auditoria, *Acustica*, Vol. 54, No. 4, (1984).

[9] M.R. Schroeder, "Modulation Transfer Functions: Definition and Measurement", *Acustica*, Vol. 49, pp 179-182 (1981).

[10] T. Houtgast, H.J.M. Steeneken, "A Review of the MTF Concept in Room Acoustics and its Use for Estimating ...", *J. Acoust. Soc. Am.*, Vol. 77, No. 3, pp 1069-1077, (1985).

[11] T. Houtgast, et Al, "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function: I. General Room Acoustics", *Acoustica*, Vol. 46, pp 60-73, (1981).

[12] D.B. Keele, "Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the Use of Time Delay Spectrometry", *AES 6th Int. Conf.*, pp 58-68, (1988 May).

[13] K. Hojberg, "STI Measurements using a Dual Channel Analyzer", *Proc. Inst. Acoust.*, (copy received from Bruel and Kjaer, London).

[14] V.M.A. Peutz, "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," *J. Audio Eng. Soc.*, Vol. 19, No. 11, pp. 915-919, (1971).

[15] C.P. Davis, "Measurement of % ALcons", *J. Audio Eng. Soc.*, Vol. 34, No. 11, pp. 905-909, (1986).

[16] "American National Standard Methods for the Calculation of the Articulation Index", S3.5-1969, *American National Standards Institute* Inc, (1969 Jan).

[17] C.M. Harris, *"Handbook of Noise Control"*, pp 14-1 to 14-9, McGraw-Hill, 2nd Edition, (1979).

[18] D. Davis, C. Davis, "Application of Speech Intelligibility to Sound Reinforcement", *J. Audio Eng. Soc.*, Vol. 37, No. 12, pp 1002-1019, (1989 Dec.).

[19] P.W. Barnett, P.H. Scarborough, "STI Measurements on Simulated Acoustic Environments", *Proc. Inst. Acoust.*, Vol. 11, pt.7, pp 201-211 (1989 Nov.).

[20] P. Mapp, P.Doany, "Speech Intelligibility Analysis and Measurement for a Distributed Sound System in a Reverberant Environment." *AES 87th Convention*, New York City, (1989 Oct).

[21] P. Doany, P. Mapp "Recent Advances in Speech Intelligibility and Sound System Design: Part II", *Proc. Inst. Acoust.*, Vol. 11, pt.7, pp 213-223, (1989 Nov).

[22] *"Speech Transmission Meter Type 3361"*, Instruction Manual, Bruel & Kjaer, (1986 Jan).

[23] R.A. Metkemeijer, J. Werff, "Design of Sound Systems with the Lowest Possible Signal to Noise Ratio", *AES 86th Convention*, Pre-print 2816 (N-4), Hamburg, (1989 March).

[24] N. Abdulrahim, P. Doany, "Recent Advances in Speech Intelligibility and Sound System Design: Part I", *Proc. Inst. Acoust.*, Vol. 11, pt.7, pp 189-200, (1989 Nov).

APPENDIX I

This Appendix explains the curves plotted in Figure 1. The first set of curves (a), (b), and (c) are plotted against reverberation time T (or early decay time EDT, for exponential decay), and the second set of curves (x), (y), and (z) are plotted against ALcons (%), as follows:

(a) STI Method: STI vs T
This is denoted curve (a) in Figure 1, and is based upon the expressions for operation beyond the critical distance, with exponential decay, as described earlier in Eqs (4-6).

(b) Peutz: Eq-STI vs T
This is denoted curve (b) in Figure 1, and is based upon Peutz's [14] expression of ALcons = 9T%, converted to Eq-STI as derived by Becker [15], so that:

$$\text{Eq-STI} = -0.1845 \ln(9T) + 0.9482. \qquad \text{(i)}$$

(c) Modified - S/N: Eq-S/N vs T
This is denoted curve (c) in Figure 1, and is based upon a 70ms integration time (τ), so that:

$$\text{Eq-S/N} = 10 \log[\exp(13.8\,\tau/T)-1] \text{ dB}. \qquad \text{(ii)}$$

(x) STI to ALcons Equivalence : Eq-STI vs ALcons
This is denoted curve (x) in Figure 1, and is based on Becker's [15] expression:

$$\text{Eq-STI} = -0.1845 \ln(\text{ALcons}) + 0.9482 \qquad \text{(iii)}$$

(y) AI to ALcons Equivalence : AI vs ALcons (or Eq-S/N vs ALcons)
This is denoted curve (y) in Figure 1, and is based upon an equivalent S/N value assumed to be equal over the entire band, which explains the scale on the AI axis as follows:

$$\text{AI} = .0333\,[\text{S/N} + 12]. \qquad \text{(iv)}$$

Curve (y) is then transcribed from ANSI S3.5 [16] for the case of rhyme tests; a slight shift to the left is needed for the case of mono-syllabic multiple-choice tests [17]. This curve is included for reference only, as it strictly applies for the case of S/N. It does however indicate the potential of an Eq-S/N being considered, based upon a 70-80ms integration time for values of ALcons in the range 5-15%.

(z) Eq-S/N to ALcons Equivalence : Eq-S/N vs ALcons
This is denoted curve (z) in Figure 1, and is essentially similar to curve (c), but with an ALcons of 9T% based upon Peutz as follows:

$$\text{Eq-S/N} = 10 \log[\exp(124.2\,\tau/\text{ALcons})-1] \qquad \text{(v)}$$

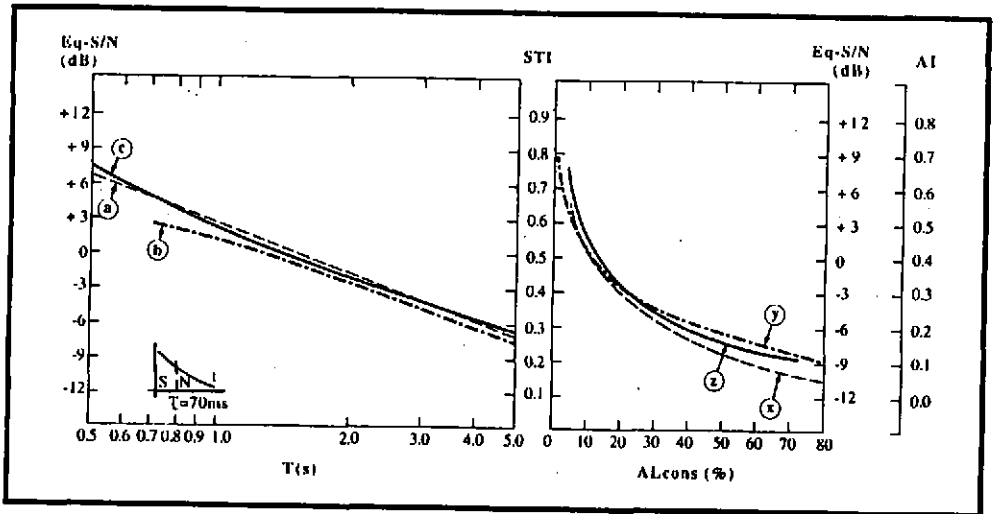

Figure 1 Summary of intelligibility parameters by different methods, for the case of pure reverberation (exponential decay): first set (a,b,c) gives STI, Eq-STI (Peutz) and Eq-S/N (70 ms) against reverberation time T, respectively; second set (x,y,z) gives Eq-STI (Becker), AI (or S/N) and Eq-S/N (70 ms) against % ALcons, respectively. Refer to Appendix I for the expressions.

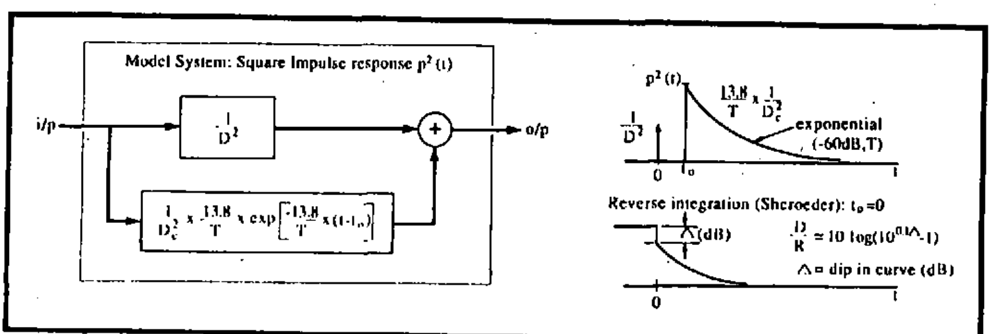

Figure 2 Ideal model for operation in a reverberant environment (exponential decay, with a direct field contribution.

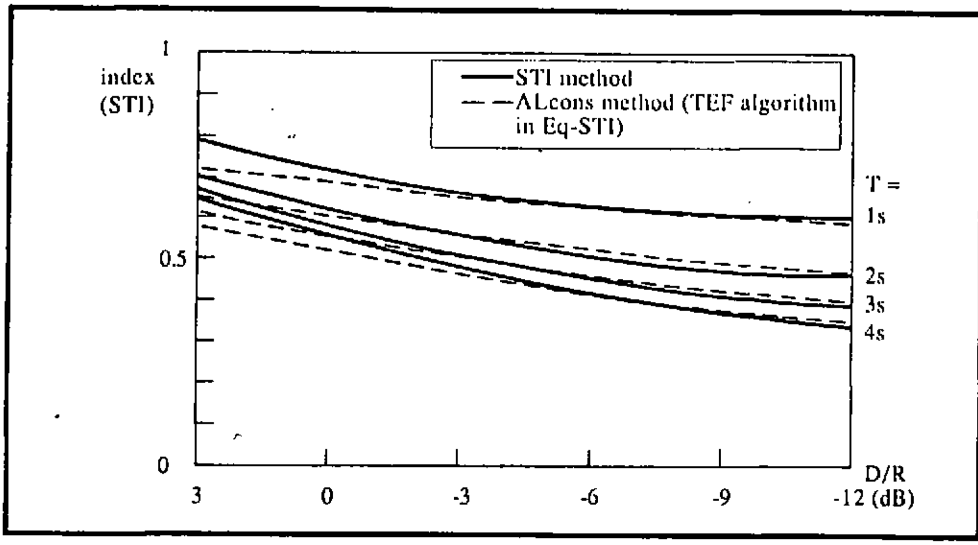

Figure 3a STI and Eq-STI (from TEF/ALcons algorithm) against D/R (dB); reverberation time is the parameter, with S/N>25 dB.

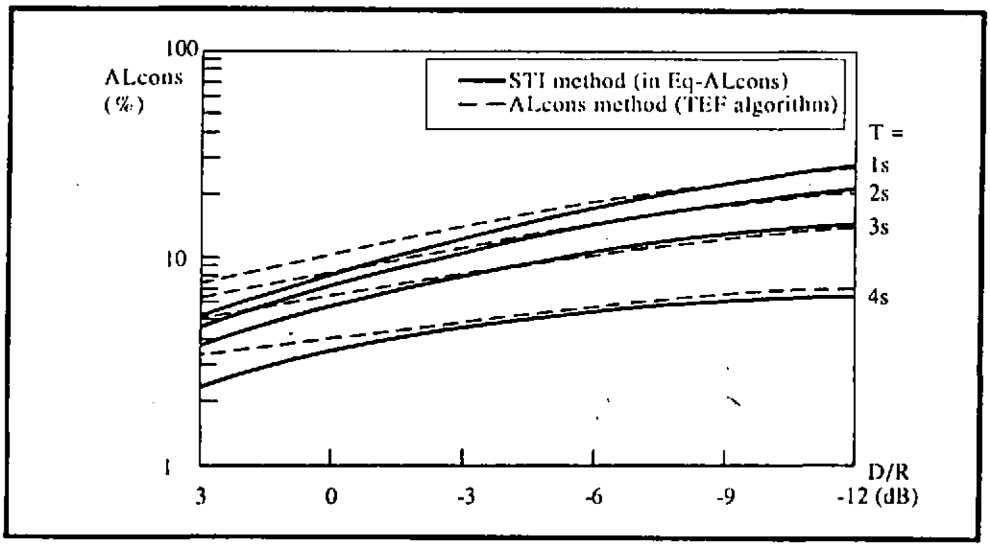

Figure 3b Eq-ALcons (from STI) and ALcons (TEF algorithm) against D/R (dB); reverberation time is the parameter, with S/N>25 dB.

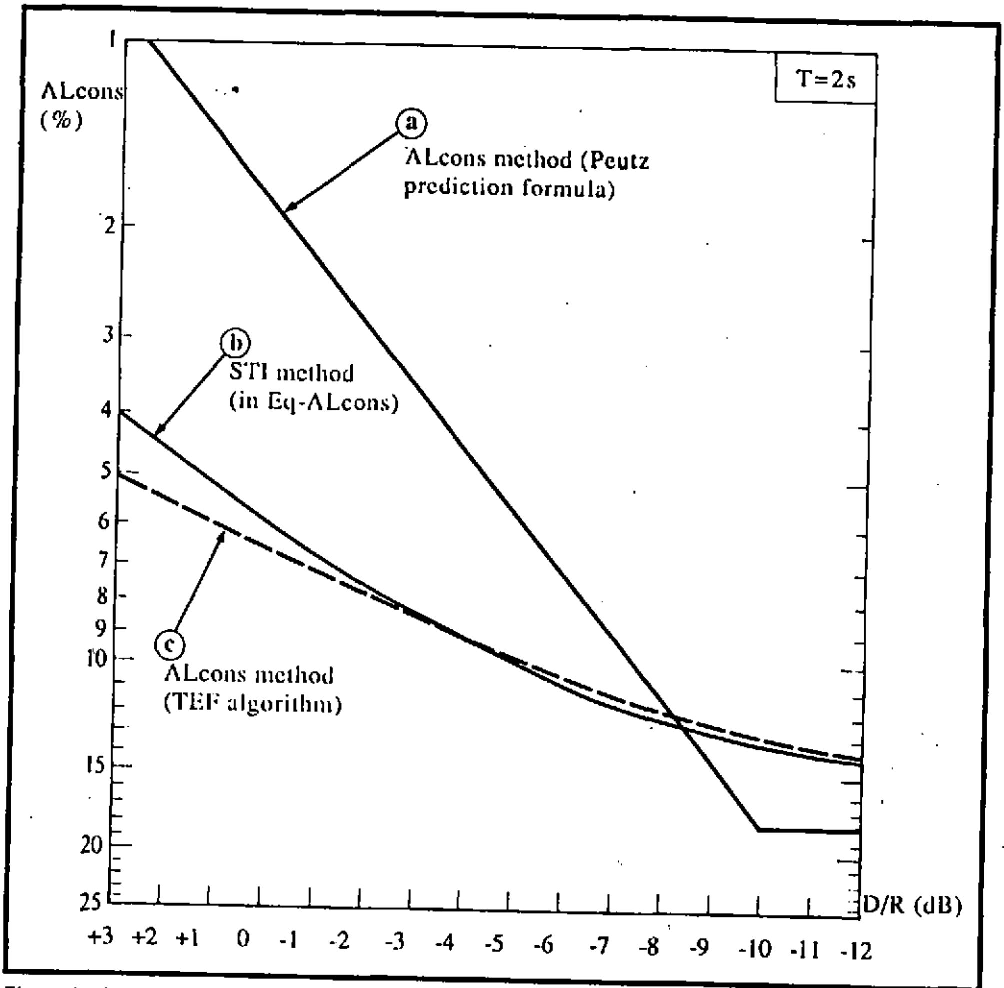

Figure 4 Percentage ALcons against D/R (dB) for T=2s (S/N>25 dB) for three methods: (a) is ALcons based on the Peutz prediction formula; (b) is STI method as outlined in this paper (Eqs. 8-10), transformed to Eq-ALcons; (c) is ALcons method based on the TEF algorithm. Curves (b) and (c) are also given in Figure 3b, but with an inverted ALcons scale.

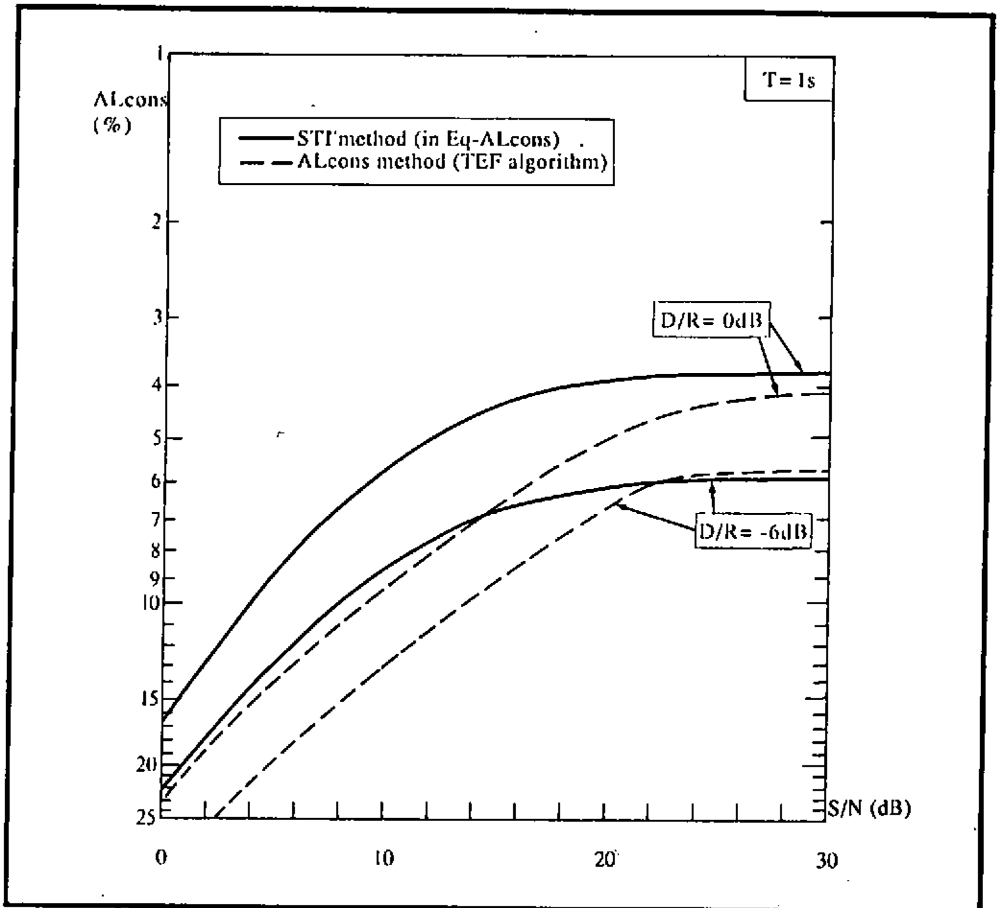

Figure 5 Percentage ALcons against S/N (dB) at two values of D/R (0 & -6 dB) for the STI method (in Equivalent ALcons) and the ALcons method (TEF algorithm).

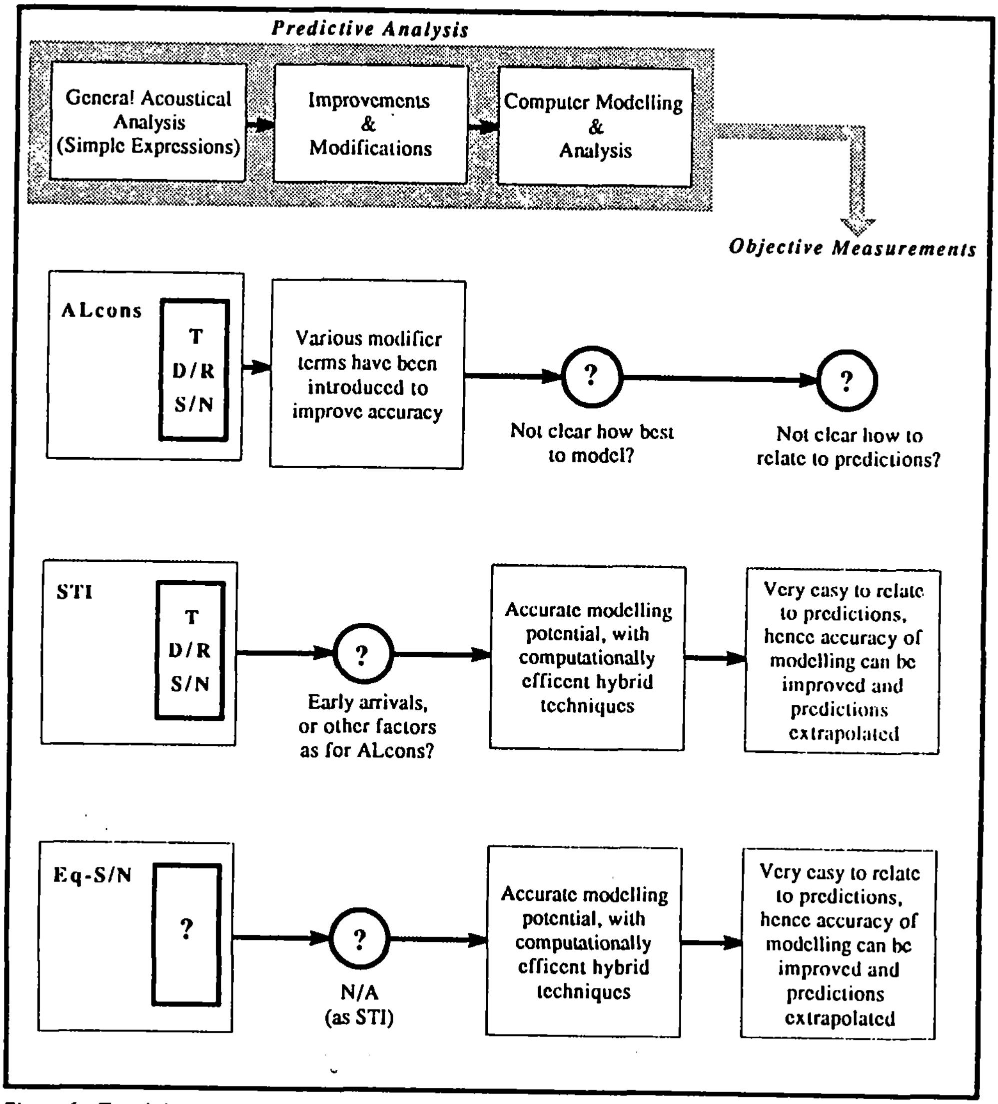

Figure 6 Trends in Predictive Analysis, Objective Measurements and their inter-relationships for ALcons, STI and S/N methods.