Proceedings of the Institute of Acoustics

A COMPUTER BASED SYSTEM FOR ASSESSING SPEECH INTELLIGIBILITY.

D.S.Brown, BSc(Eng), A.C.G.I.

Cirrus Research plc, Acoustic House, Hunmanby, United Kingdom.

1. INTRODUCTION.

Traditional building acoustics methods for determining measures of a room's parameters do not give a complete picture of the acoustic environment. A proven laptop computer based solution, part of the ARIA acoustic measurement system, uses a stationary pseudo-random white noise source to determine speech transmission indices. These indices will be compared and contrasted with other measured parameters such as clarity, definition, and early decay time, thereby illustrating the desired flexibility of such a measurement system.

2. REVERBERATION TIME - THE TRADITIONAL MEASURE.

The most common building acoustics measure is reverberation time, R_{T60} - the time interval for sound levels to linearly decay by 60 dB when the room is excited by a sound source. Generally, reverberation time is calculated as double the time taken to decay from -5 to -35 dB.

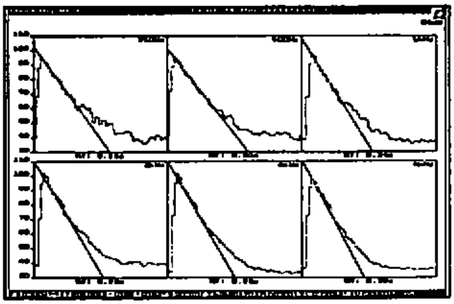

Figure 1 : Octave Band Reverberation Time Curves.

Such a simple measure whilst useful can only give an impression of the liveliness or reverberative nature of a room - for normal speech to avoid blurring R_{T60} should be between 1 to 1.3 seconds. However, for highly reverberant rooms the reverberating part of the echogram may combine with the direct sound to yield a fallacious result.

3. MEASURES OF INTELLIGIBILITY

Intelligibility criterion are essential to truly characterise a room's acoustic performance when sound is emitted in a room either directly or via an amplification system. Then, by virtue of a series of reflections and reverberations, the listener's ear receives a less than perfect impression of the original signal. With a "very live" or highly reverberant room then the combined effect of successive reverberations can be such to render sound unclear.

Hence different objective parameters have been suggested to qualify a room's intelligibility.

3.1 Early Decay Time, Clarity and Definition.

Early decay time (EDT) is calculated on the first 10 dB of decay and can be argued to represent the ear's response to the acoustic decay of large signal levels. When compared with R_{T60}, EDT will give a measure of the early energy decay. Since EDT accounts for the direct sound and its first reflection it is close to the subjective judgement of the listener.

Another obvious parameter to use is a ratio of the acoustic energy before a given time period to the energy after that time. Essentially, the principle is the same - energy before a period is considered useful and the energy after that time is considered either secondary or a nuisance. Several times : 33, 35, 50, 80 and 95 milliseconds and two indices : clarity and definition have been suggested for this purpose - D_{50}, the 50 milliseconds definition in conjunction with C_{80} the 80 milliseconds clarity for music or with C_{50} for speech with have been suggested

Clarity is

$$C_{80} = 10\log\frac{\int_0^{80ms} h^2(t)\,dt}{\int_{80ms}^{\infty} h^2(t)\,dt} = 10\log\frac{E_{80}}{E_{rev}}$$

Clarity will vary throughout an auditorium, and for music suggested limits are from -2 to 2 dB. In practice values larger variations may be found and typical C_{80} clarity ranges are given below (1) :

Venue.	Lowest Value.	Average Value.	Highest Value.
Gross Musikhereinsalle, Vienna.	-5.30	-2.90	-0.30
Salle Pleyel, Paris.	-3.37	-0.70	4.37
Auditorium M. Ravel, Lyon.	-3.07	-0.13	2.88

Typical Auditoria C_{80} Clarity Values.

Whereas clarity is a "before to after" ratio, an alternative has been suggested called definition, which is the ratio of the energy before a given time to the total energy, with the reference time

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

normally being chosen as 50 milliseconds.

$$D_{50} = \frac{\int_0^{50ms} h^2(t)\, dt}{\int_0^{\infty} h^2(t)\, dt} = \frac{E_{50}}{E_{tot}}$$

Whilst clarity and definition indices undoubtably enhance the assessment of a room's intelligibility, they are often unreliable and do not account for background noise.

3.2 Lochner Burger Signal to Noise Ratio.

Lochner and Burger suggested an intelligibility ratio be used when the reverberation dominates the original signal. This is based upon the premise that reverberated sound energy acts as a parasitic noise source that perturbs the understanding of words transmitted on the direct path and its first reflection. Up to 35 milliseconds, all the sound energy is considered useful, between 35 and 95 milliseconds it has a decreasing contribution and thereafter, the sound energy does not make any contribution. The higher the ratio of the pre 95 milliseconds energy to the energy afterwards then the more intelligible the speech.

$$Lochner\text{-}Burger\ Signal\ Noise\ Ratio = 10\log\frac{\int_0^{95ms} a(t)\, h^2(t)\, .\, dt}{\int_{95ms}^{\infty} h^2(t)\, .\, dt}$$

3.3 RASTI and STI.

STI is an objective criteria used to characterise speech intelligibility, by accounting for all possible causes of speech alteration other than non-linearity.

Any alteration in signal modulation can be expressed as a signal to noise ratio - the modulation transfer function where the numerator is the Fourier transform of the squared impulse response.

$$m(F) = \frac{\int_0^{\infty} h(t).e^{2j\Pi Ft}\, dt}{\int_0^{\infty} h^2(t).dt} \cdot \frac{I_{Speech}}{I_{Speech}+I_{Noise}}$$

Each octave's signal to noise ratio can then be calculated as -

$$S/N_k(F) = 10\log\frac{m(F)}{1-m(F)}$$

and an average ratio is calculated after limiting the each frequency band's ratio to $\pm$ 15 dB as

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

$$\overline{S/N_k} = \frac{1}{n}\sum_F S/N_k(F)$$

Speech transmission indices are then calculated for each frequency band as below :

$$\overline{TI_k} = \frac{\overline{S/N_k} + 15}{30}$$

Finally, STI, the Speech Transmission Index is calculated as a weighted sum of each frequency band's transmission indices.

$$STI = \frac{\sum_F (W_k \cdot TI_k)}{\sum_F W_k}$$

STI can be used to determine the critical distance, d_c, at which point intelligibility deteriorates. Typically, STI slowly decreases with distance until at d_c the value becomes almost constant as reflected sound dominates.

For a typical conference room with α between 0.1 and 0.4 then d_c normally lies between 15 to 20 metres.

The RASTI or Rapid Speech Transmission Index calculates similar objective measures of speech clarity using a similar methodology as STI but is limited to the 500 Hz and 2 kHz octave bands - these being the pertinent to normal speech. RASTI also limits all signal to noise ratios to $\pm$ 15 dB and the index is calculated as :

$$RASTI = \frac{(\overline{X_i} + 15)}{30}$$

Since this system has the capability to calculate both RASTI and STI then this facilitates comparisons to be made with measurements from other dedicated systems with less computational power and hence only producing RASTI data.

3.4 A Measure of Balance.

Whilst all the above indices give valuable building acoustics information, there is another measure most useful in sound reproduction situations - this is the energy in each octave band expressed as ratio to that contained in the 1 kHz band. Such an index is useful giving the measurer a feel of the "colour" of the room and it's orchestral balance (bass, middle, and treble). The experienced sound engineer is then able to see which frequencies need most amplification.

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

4. IMPLEMENTATION OF AN INTELLIGIBILITY ANALYSER USING THE ARIA SYSTEM.

4.1 System Methodology.

ARIA uses digital sampling techniques based on Nyquist's theorem to truly represent an acoustic input. The building acoustics measurement system, as shown below, consists of a high performance digital signal processor card in a host computer with a microphone input and generating a white noise signal via an output channel. The dBIMPULS software module stores the raw signal file, treats the received signal to derive the echogram and generates data such as RASTI and STI coefficients and generates measurement reports.

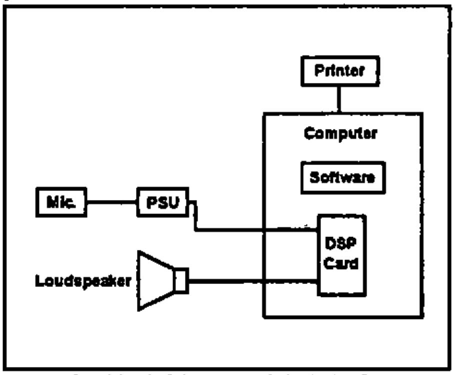

Figure 2 : Block Diagram of ARIA's Computer Based Intelligibility Analysis and Reporting System.

4.2 Computer Requirements.

A highly transportable system must have relatively easy to fulfill computer requirements. The software operates from hard disk under the globally acceptable MS-DOS operating system and uses an 80286 or better processor with a basic memory requirement of 512K of useable RAM. The digital processor card provides the main computer constraints : a full size 16 bit expansion card is needed and the hard disk access speed should be less than 28 milliseconds. Such requirements are easily satisfied by laptop computers such as the Toshiba T3200SXC and T6400 units.

4.3 System Philosophy.

The concept of a modular computer based acoustic measurement system was proposed by Luquet and Rozwadowski (2) and resulted from earlier production of a data storing sound level meter (3). From its conception, the ARIA system was not conceived to be another acoustic instrument, but to be a totally new type of measurement acoustics. The system's philosophy is not to focus on the measurement, but to consider the complete measurement and reporting process.

With this in mind, the ARIA system uses the host computer's calculating power and attendant

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

peripherals to completely revise the measurement process : wherever possible, raw data is stored and analytical decisions can be made after the raw data has been collected. Data can be reanalysed to allow multiple scenarios to be investigated from a single datafile, thus saving the acoustic practitioner valuable time and post measurement dissonance. The use of a report generator and standard output devices such as laser-jet printers result in an integrated measurement, analysis and reporting system (4).

4.4 The Computerised Building Acoustics Analyser.

Outbox processing techniques used in the early spectral analysis and L_{eq} modules are ideally suited to many building and architectural acoustics measurements. The processor card contains fast fourier transform routines that allow the stored signal files to be resolved in the frequency domain into octave, one third octave, or narrow band spectra. Initially, a reverberation time add-on module allowed measurement of octave and one-third octave R_{T60}'s and determination of isolation levels L_{nw} and R'_{w}.

However, to address the problem of intelligibility, a new module, dBIMPULS, was developed with a French centre of excellence, the Centre Scientifique et Technique du Batiment (C.S.T.B.) in Grenoble.

dBIMPULS determines a building's impulsive response to an internally generated pseudo-random white noise burst. By using a wideband noise signal the limitations of traditional sources such as gunshots namely non-linearity and poor dynamic range are avoided.

The user specifies a range of initial parameters such as storage filename, sequence length, and sampling frequency. These parameters result in the production of a customised white noise burst alongwith optional automatic gain adjustment and subsequent recording of the returning signal.

```
System   Measurement   Analysis   Configuration                12:22:58

                        Acquisition setting

Output file name ( without ext. ) : DEMO
Comments : This is a demonstration SETUP.
Sequence order : 13
Sampling frequency (kHz) : 16
Averages                 : 4
Input selection          : MICRO
Input gain               : AUTO
```

Initial Acquisition Options.

Once the returning signal is recorded on the computer's hard disk, then the user is able to view the linear trace of the room's impulse response or alternatively, this may be displayed in dB - a squared impulse response.

By performing an FFT analysis on the received signal then architectural indices such as RASTI, STI, and the Lochner-Burger signal to noise ratio may be calculated as shown in the table below or alternatively the octave band echogrammes and decay curves may be viewed as shown below the

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

results table.

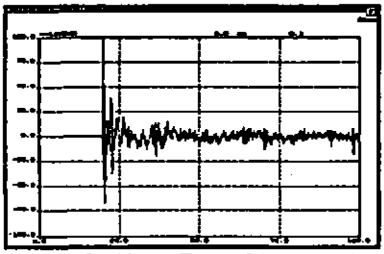

Figure 3 : Linear Trace Of A Room's Impulse Response.

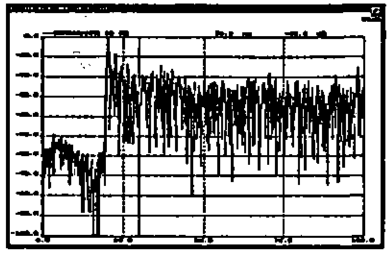

Figure 4 : The Same Room's Echogramme in dB.

D:\01DB\DEMO\CAV1.FIL
Centre between 2 Aisles.
Sequence order= 14 Sampling frequency = 16 kHz
Auto RT (*) and EDT without noise other crit. without noise
RASTI = 0.83 STI = 0.78

Octave Hz	RT60 s	EDT s	C80 dB	D50 %	S/N dB	E1000 dB
125	* 0.53	0.47	12.34	78.81	12.84	-10.9
250	* 1.07	0.44	12.92	82.35	14.22	-8.2
500	* 0.46	0.27	14.69	93.20	15.50	-1.0
1 k	* 0.66	0.38	11.99	90.05	12.77	0.0
2 k	* 0.69	0.51	10.26	87.56	11.03	2.6
4 k	* 0.62	0.50	10.45	85.44	10.90	1.8

Results Table Produced from dBIMPULS.

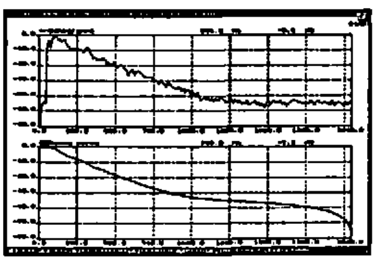

Figure 5: Echogram and Schroeder Decay Curve Produced By Analysing A Room's Impulsive Response.

As we have seen differing times have been suggested for clarity and definition indices, so a flexible system must allow the user to customise the system and its configuration menu is shown overleaf. Hence, a highly flexible system has been produced which simultaneously measures clarity, definition, Lochner Burger signal to noise ratio, RASTI and STI indices.

COMPUTER BASED SPEECH INTELLIGIBILITY ANALYSIS.

```
                         Configuration

FIRST OCTAVE BAND for criteria display (Hz) : 125
LAST  OCTAVE BAND for criteria display (Hz) : 4 k
Clarity and definition INDEX CAPTURE        : No
Default CLARITY INDEX                  (ms) : 80
Default DEFINITION INDEX               (ms) : 50
Criteria CALCULATION with NOISE SUBTRACTION : Yes
Max SCALE on 0 dB                           : Yes
S/N ratios for STI & RASTI calculation      : Infinite
```

4. MEASUREMENT ACCURACY.

While it is relatively easy to develop an indicative system, all ARIA system modules were designed to provide precision solutions to known measurement problems. This has resulted in the component specifications far in excess of those normally associated with computer based instrumentation and a resultant associated cost. Such a philosophy was vindicated when ARIA became the world's first computer based acoustic measurement system to be verified as a type 1 sound level meter (5). Whilst no global standard yet exists for FFT analysis, the FFT module has shown excellent agreement with classical analysers in independent comparative tests (6).

5. PREDICTIVE SOFTWARE.

An obvious development of this technology is in the production of predictive software and the EPIDAURE program (7) has been produced, again in collaboration with the C.S.T.B., to enable STI and similar parameters to be predicted and mapped. This is a subject worthy of further discussion but is unfortunately outside the scope of this paper.

6. CONCLUSION.

A precision computer based intelligibility analysis and reporting system has been successfully implemented in a laptop computer. This gives an easy to use, user-configurable measurement system with the ability to reanalyse data, and allows impressive presentation from a lightweight and portable solution to intelligibility assessment.

References

1 Jullien J.P. "Acoustique des Salles" CNET Lannion, 1982 p 19.

2 Luquet P, and Rozwadowski A. " PC Based Integrated Acquisition and Data Processing System " Proc Internoise pp 241-244 Avignon, France August 1988

3 Wallis A D, and Luquet P. "Computer Acquisition of Large Data Sets" Proc Internoise pp 1423-1426 Beijing, China. September 1987

4 Brown D S, and Ethore C. " ARIA - A Comprehensive Acoustic Measurement System For Personal Computers." Proc Euronoise pp 777-784 London September 1992

5 Ministere de l'Industrie et de l'Amenagement du Territoire. " Decision d'Approbation de Modele no. 90.2.01.931.1.1 du 10 Dec 1990. Paris 1990.

6 Martin C. and Roland J. "Logiciel ARIA implante sur micro ordinateur COMPAQ III equipe de carte AU 22." CSTB Grenoble Report No 2/89/110 Mai 1989.

7 Brown D S, and Thiebaut L. "The Computer Ray Tracing Method Using ARIA" IOA Reproduced Sound 1991.