Proceedings of the Institute of Acoustics

OVERVIEW OF SPEECH INTELLIGIBILITY

P.W. Barnett AMS Acoustics, Rayleigh House, Bush Hill Park, EN1 2QB.

1. INTRODUCTION

The science and in particular the application of the technology of speech intelligibility is relatively new - a product of this Century. Purists or those with a long memory might remind us of the reported attempts by Greeks and Romans to improve communication in large spaces but in my opinion this relates rather more to loudness .

A considerable debt must be owed to pioneers of the telephone industry on or around the turn of the Century who laid down many of the principles in use today. Interest by those in the telephone industry has been joined by others over the last 100 years and include Military communications, education establishments, auditoria together with public address systems. Recently we need to add those concerned with speech synthesis, speech-to-text and text-to-speech systems.

Speech intelligibility is now embodied in quasi legislation following Hillsborough, Bradford and Kings Cross disasters and subsequent Inquiries where reduced speech intelligibility was either directly cited or implied. Three major Standards now include a direct reference to speech intelligibility and from an acoustical standpoint BSEN 60849 represents a milestone since it includes the CIS scale as shown in fig. 1.

The CIS system was a product of an Institute Of Acoustics Working Party .

This overview Paper is intended to provide an insight to the factors that affect speech intelligibility and to provide examples as appropriate. To facilitate this I propose to use a base model which includes most of the relevant factors.

2. BASE MODEL

Knudsen proffered (1929) that

Percent Articulation = $A\% = 96\ K_e K_r K_n K_s$ ....(i)

where: $A\%$ = % Articulation.
K_e = reduction factor owing to inadequate loudness
K_r = deduction (or distortion?) factor relating to reverberation
K_n = reduction factor owing to noise
K_s = reduction factor owing to the shape of the room.

All reduction factors are in the range 0-1.

The factor 96 Knudsen suggested represented the maximum since 100% was not attainable.

From equation (i) we can invoke the concept that no factor can exceed unity and hence it is not possible to improve one factor to account for the deficiency of another. Whilst this might not be strictly true today , the principle is largely robust and is reflected in Speech Transmission Index (STI) after Houtgast et al which provides us with one solution that:

$$m(F) = \frac{1}{\sqrt{1+\left(\frac{2\pi FT}{13.8}\right)^2}} \cdot \frac{1}{1 = 10^{\frac{-S/N}{10}}}$$

where: m(F) = modulation reduction factor
F = modulation Frequency (Hz)
T = Reverberation Time (secs.)
S/N = Signal-to-Noise ratio (dB).

In this instance the modulation reduction factor is the product of the two multiplicands, the first of which relates to the space and the second relates to the S/N ratio.

It can be seen by inspection that each multiplicand lies in the range 0-1 and hence the deficiency rule applies.

If we return to Knudsen model:

$$A\% = 96\ K_e K_r K_n K_s$$

we can consolidate some reduction factors, add two additional factors and re-suffix the parameters in line with current practice.

vis:

$$A\% = P_t P_l K_{S/N} K_{DIR} K_T K_S F_p$$

where: P_t = talker Proficiency

P_l = listener Proficiency

$K_{S/N}$ = reduction factor due to poor signal-to-noise

K_{DIR} = reduction factor due to poor direct-to-reverberant ratio

K_T = reduction factor due to Reverberation Time

K_S = Reduction factor due to the room shape

F_p = Factor for signal processing.

It would be clearly possible to sub-divide the some of the reduction factors and to add further multiplicands but the foregoing probably represents the majority of cases.

The following illustrations (mostly unpublished) are derived from our work and researches over the past 10 years.

3. ILLUSTRATIONS

3.1 Illustration 1 - Talker Proficiency

Proficiency factor sometimes referred to as practice factor, relates to the proficiency of a talker-listener combination. This factor which lies in the range 0-1 can be expressed as the product of the talker and listener proficiency factors i.e. $P = P_t . P_l$.

For most intelligibility testing using subjective methods, the objective is to obtain a proficiency factor of unity i.e. $P_t = 1$ and $P_l = 1$.

Talker proficiency depends upon training and the ability to articulate correctly.

Fig. 2 shows data collected under the same conditions in the same space (church hall) at the same measurement position for three different talkers:

1. Female - Practised talker - perfect diction.
2. Male - actor - excellent diction.
3. Female - teacher - good diction.

It can be seen that the difference increases in the presence of noise. Furthermore P_t is a function of S/N.

If we accept that $P_t = \frac{\%C_x}{\%C_{pd}}$

where: $\%C_x$ = percent correct for talker under test and $\%C_{pd}$ = percent correct for talker with perfect diction i.e. $P_t = 1$.

then P_{tx} is clearly a function of S/N.

Fig. 3 shows a graph of deduced talker proficiency for the data given in fig. 2.

3.2 Illustration 2 - Talker Proficiency for Text-to-Speech System

In this experiment the Word Score material was produced on a text-to-speech system. The experiment involved the preparation of several word Score sets on a text-to-speech signal and to input the lists to a reverberant space via a PA system. Comparative tests were also carried out using lists prepared in the usual way.

Fig. 4 shows some of the results obtained. It can be seen that the text-to-speech signal fared particularly badly.

The point of this illustration is that those concerned with the text-to-speech system had spent considerable effort and energy in ensuring that the speech sounded natural and that each word was correctly emphasised, if this technically is to be employed in PA/VA environments then it should be possible to process the speech to make it more intelligible.

3.3 Illustration 3 - Listener Proficiency

There are many examples of this, since the major problem with subjective testing is to ensure that listener proficiency approaches unity. In fact due to the recognition aspect, it is possible to have advanced listener proficiency of greater than unity and for this reason listening crews are graded .

Fig. 5 shows the difference in intelligibility for a group of listeners with $P_l \approx 1$ and a listener with presbycusis (age 65) .

It can be seen that the difference increases with decaying S/N ratio.

Fig. 6 shows the calculated listener proficiency with S/N as the independent variable.

3.4 Illustration 4 - Effect of S/N

Previous illustrations have been with S/N as the independent variable and there is little more to say except that from our work we have deduced that a third order polynomial represents the best-fit curve as shown below:

$$\%I = Ax^3 + Bx^2 + Cx + D$$

where: I = % Words correct
X = S/N dB
A lies in the range 0.002 to 0.01
B lies in the range -0.10 to -0.35
C lies in the he range 1.0 to 6.0
D lies in the range 0 - 100.

Each polynomial is space and position dependent.

Fig. 7 shows the data points obtained for 5 No. positions in a large reverberant multi-purpose sports hall (volume = 3500m^3, RT $\approx$3.0 secs.). Data is for female voice [$P_l \approx 1$] CVC Words. Increasing position numbers are for increasing distance from an omni-directional source.

3.5 Illustration 5 - Effect of D/R

Fig. 7 (illustration 4) provided the data points for 2m, 4m, 8m, 16 and 32m, each position taken from the omni-directional source.

From measurements made at the time the direct-to-reverberant ratio at the 4m position was measured as 1.0dB, D/R at all other positions were deduced in the normal way.

Fig. 8 shows the data given in fig. 7 replotted at various S/N with D/R as the independent variable.

It can be seen that the reduction in intelligibility is less with decreasing D/R ratio than for reducing S/N.

For the space in question the intelligibility at a given point can be deduced from:

$$\%I = -0.01W^2 + (-0.04x + 1.4)W + (-0.10x^2 + 4.0x + 41)$$

where: I = %Words scored correctly
W = D/R (dB)
x = S/N (dB).

The above expression is useful in determining the %I at various salient positions i.e. W=0 and/or x = 0.

The data should not be extrapolated beyond

$-20 < D/R < 10$dB.
$0 < S/N < 20$db.

3.6 Illustration 6 - Effect of RT

The effect of RT on speech intelligibility is complex since reverberation time affects direct-to-reverberant ratio. To simplify the issue, fig. 9 shows RT as the independent variable for two comparable spaces with S/N set at >25dB and D/R set at 5dB, 0dB, -5dB and -10dB. It can be seen that the change is significant with RT as the independent variable but with D/R ratio held constant but that as the direct-to-reverberant ratio deteriorates, the effect of RT reduces.

3.7 Illustration 7 - Effect of Room Shape

Knudsen from his Paper suggested that K_s for rectangular rooms circa 25m x 40m x 10m would be in the region 1 but that for other shaped rooms it might reduce to as little as 0.9.

The difficulty in assessing this factor is patent. However we were fortunate to carry out measurements in a small church hall and the Royal Festival Hall. Although the spaces differed greatly in size, their reverberation times were comparable.

The table below gives a direct comparison of the spaces:

Parameter	Space	
	Church Hall	Royal Festival Hall
Volume	$717m^3$	$21950m^3$
RT_c1kHz	1.5 sec.	1.5 sec.*
mfp	6.2m	18m
*Assisted Resonance System off		

Fig. 10 shows the data obtained in each space presented with D/R as the independent variable. It can be seen that the results for the Royal Festival Hall are much lower than those for the small church hall even though the measurement positions were acoustically comparable. Furthermore the mechanism for reduced intelligibility appears, in each case, to be quite different. We are of the opinion that this is related to the mean free path and that different a type of masking is involved.

In addition to the Word Scores RASTI measurements were taken at each measurement position and this data is presented in a similar way in fig. 11.

It can be seen that the objective results are comparable which could suggest that this objective measure does not take full account in all situations.

3.8 Illustration 8 - Effect of Frequency Response

In this experiment conditions were tested at 3 No. positions in a space for various applied S/N ratios. The spectral adjustments were as follows:

1. Low Pass filter - 4kHz cut off @ 12dB/octave.
2. Low Pass filter - 2kHz cut off @ 12dB/octave.
3. Low Pass filter - 1kHz cut off @ 12dB/octave.
4. No filtration.

The results are shown in fig.12.

It can be seen that the effect is dramatic emphasising the importance of maintaining a wide bandwidth. Tests carried out with telephone systems (bandwidth limited) do not show such a dramatic effect due to the absence of 'acoustics' and generally a better signal-to-noise ratio.

3.9 Illustration 9 - The Effect of Amplitude Compression

The results given in this illustration are taken from a major research programme undertaken by AMS Acoustics in the period 1997 - 1999. The research programme involved applying amplitude compression to speech and replaying the results in a variety of spaces to compare both compressed and uncompressed speech.

Fig. 13 shows a sample of the results obtained.

It can be seen that the application of amplitude compression provides a significant improvement and indeed suggests that the fact F_p can be greater than unity.

Fig.14 illustrates the deduced proficiency. Hence this technique may be used to restore a proficiency of 0.7 to unity.

3.10 Illustration 10 - The Effect of Amplitude Compression on a Person suffering from Presypcosis

As part of our study we used listening jurors who suffered for age-induced presbycusis, the results are shown below in fig. 15.

Hence it can be seen that for a listener with $P_l > 1$ due to presbycusis the application of compression returns the listener to $P_l \approx 1$.

Finally fig. 16 shows the listener proficiency of an individual with a degree of presbycusis increase above normal (in this situation) with the application of amplitude compression.

4. CONCLUDING REMARKS

It is hoped that the foregoing has provided an insight to the factors that affect intelligibility and the degree of effect.

We have been fortunate that much of our work has necessitated the maintenance of one or more listening crews which in turn means that there has always been room for additional tests and experiments to be applied to the main work at hand.

Particularly the effect of room shape and masking mechanisms is worthy of additional attention as is the use of signal processing techniques to increase P_l above unity or to produce truly remedial situations.

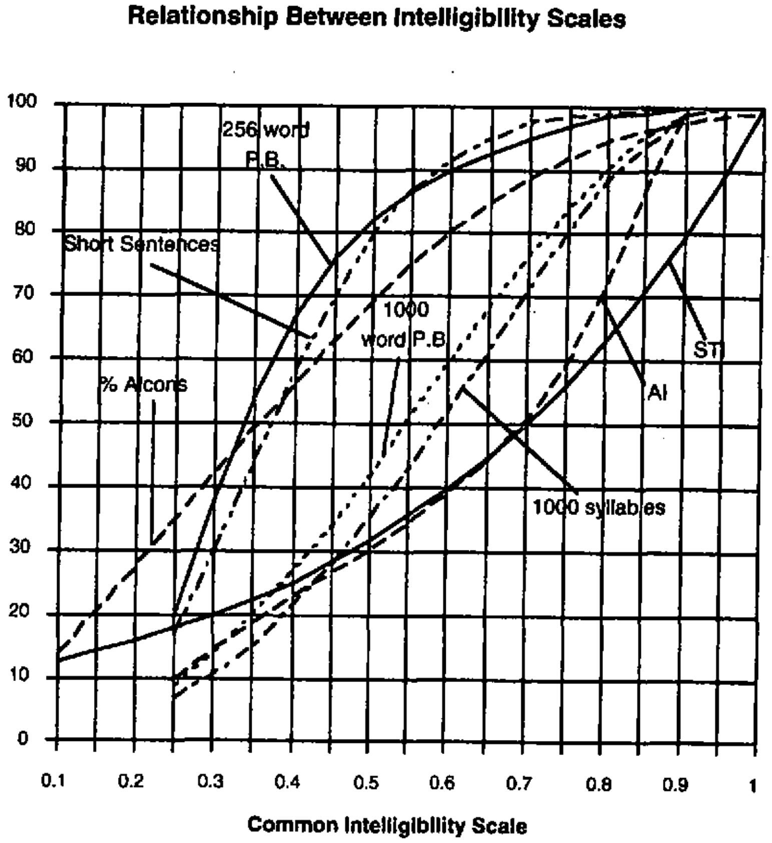

Relationship Between Intelligibility Scales
100
90
80
70
60
50
40
30
20
10
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Common Intelligibility Scale
256 word P.B.
Short Sentences
1000 word P.B
% Alcons
STI
AI
1000 syllables

Fig 1

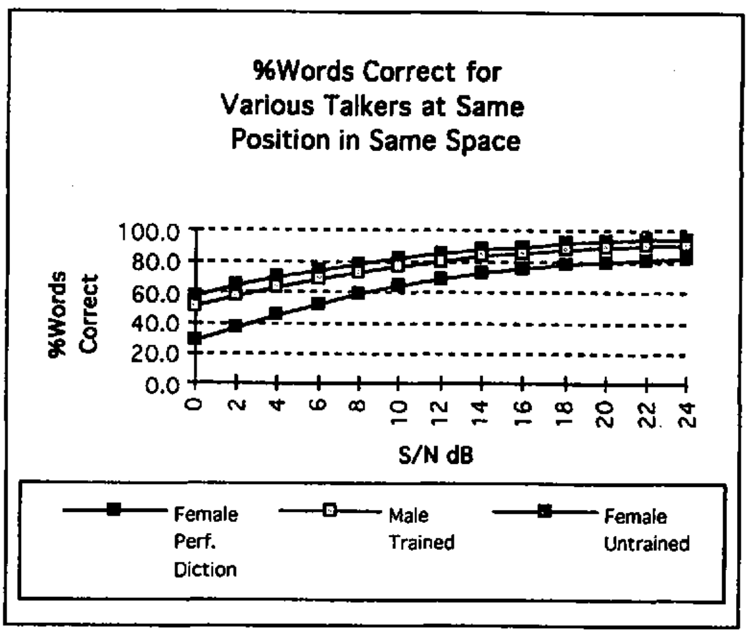

%Words Correct for Various Talkers at Same Position in Same Space
%Words Correct
100.0
80.0
60.0
40.0
20.0
0.0
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Female Perf. Diction
Male Trained
Female Untrained

Fig 2

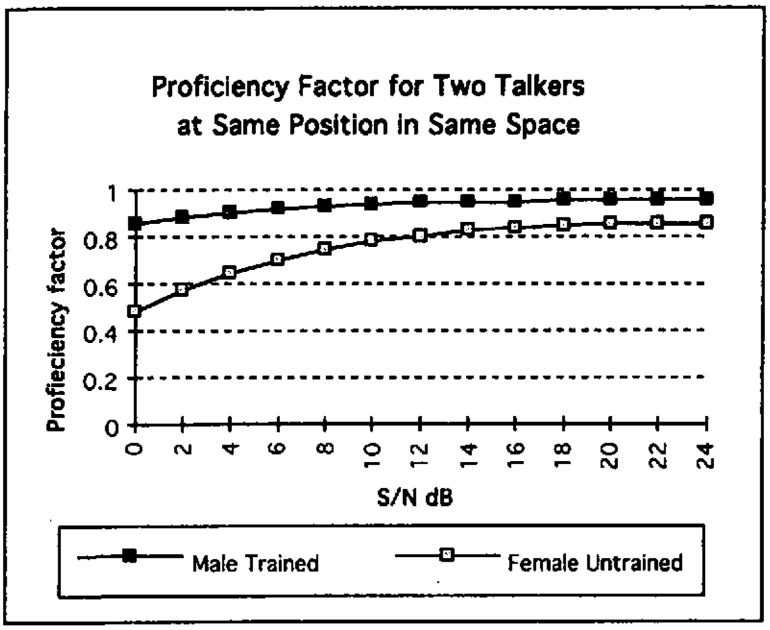

Proficiency Factor for Two Talkers
at Same Position in Same Space
Profieciency factor
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Male Trained
Female Untrained

Fig 3

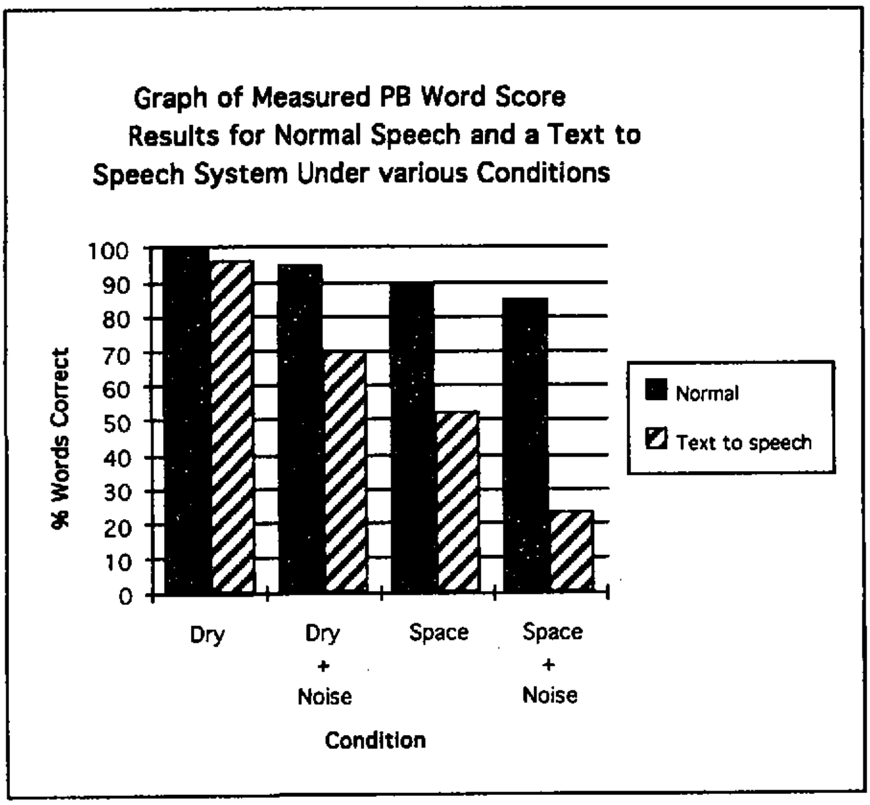
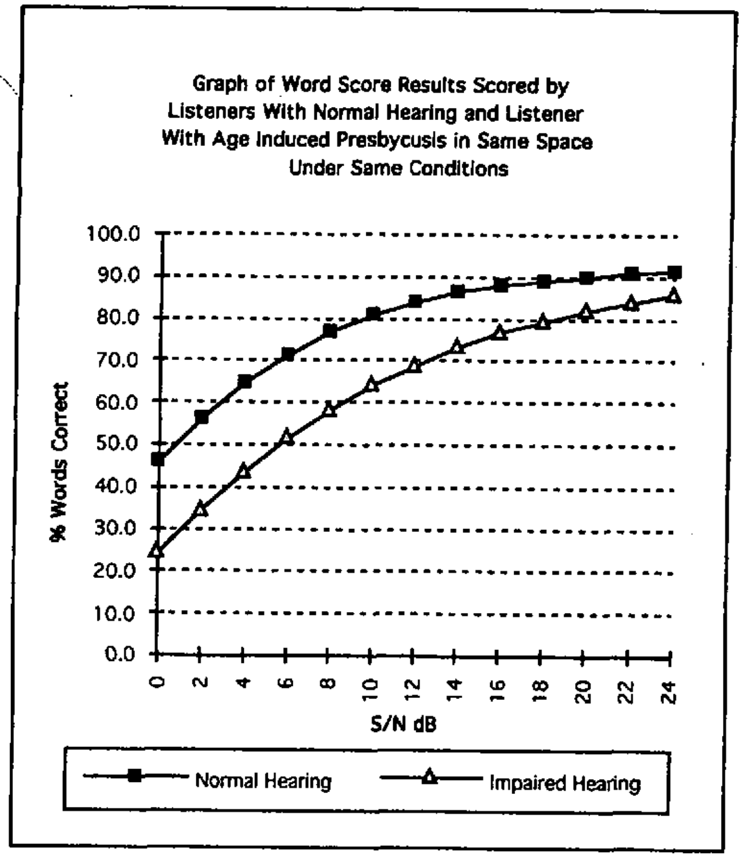
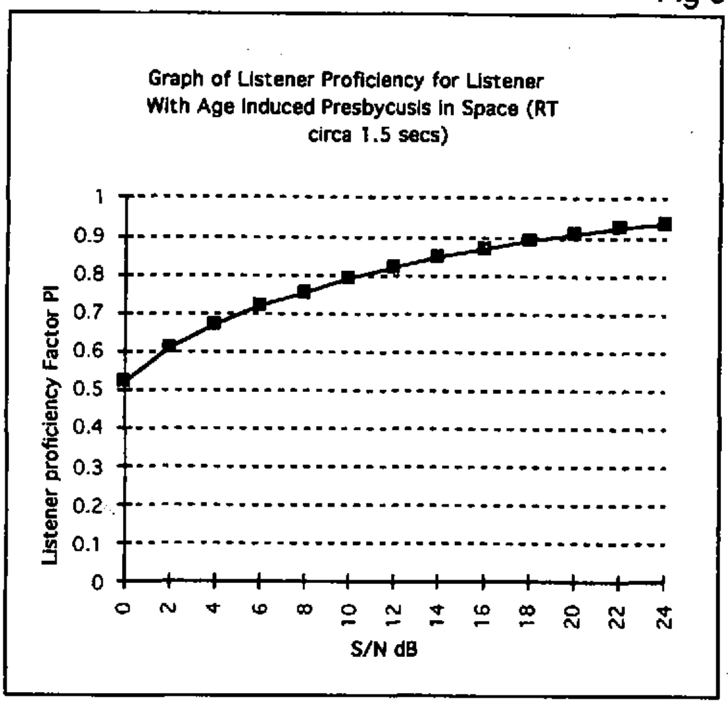

Graph of Measured PB Word Score
Results for Normal Speech and a Text to
Speech System Under various Conditions
% Words Correct
100
90
80
70
60
50
40
30
20
10
0
Dry
Dry + Noise
Space
Space + Noise
Condition
Normal
Text to speech

Fig 4



Graph of Word Score Results Scored by Listeners With Normal Hearing and Listener With Age Induced Presbycusis in Same Space Under Same Conditions
% Words Correct
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24
S/N dB
Normal Hearing
Impaired Hearing

Fig 5



Graph of Listener Proficiency for Listener With Age Induced Presbycusis in Space (RT circa 1.5 secs)
Listener proficiency Factor PI
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 2 4 6 8 10 12 14 16 18 20 22 24
S/N dB

Fig 6

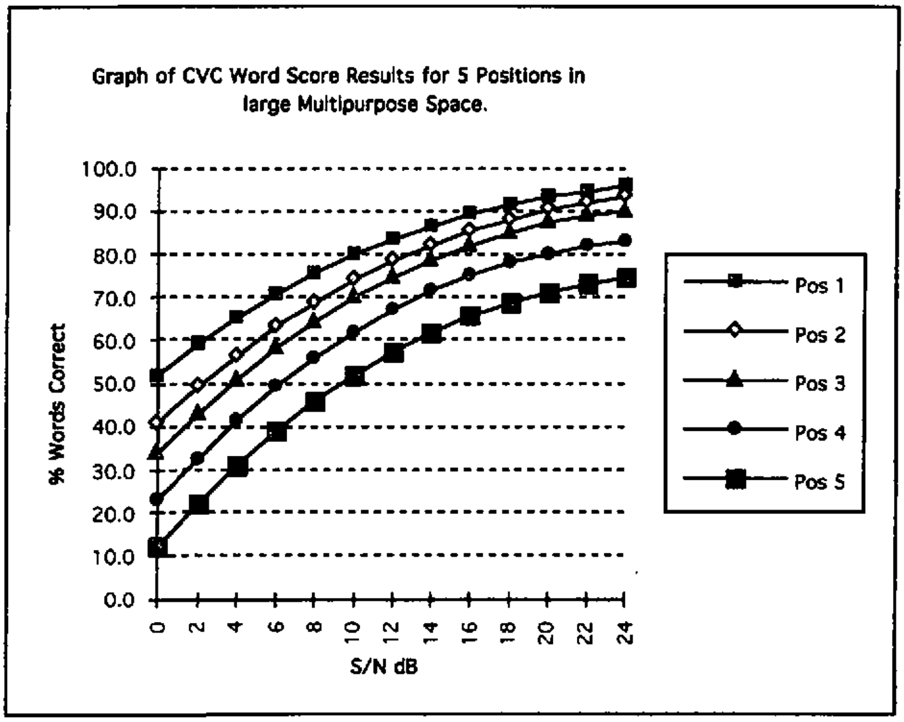

Graph of CVC Word Score Results for 5 Positions in large Multipurpose Space.
% Words Correct
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Pos 1
Pos 2
Pos 3
Pos 4
Pos 5

Fig 7

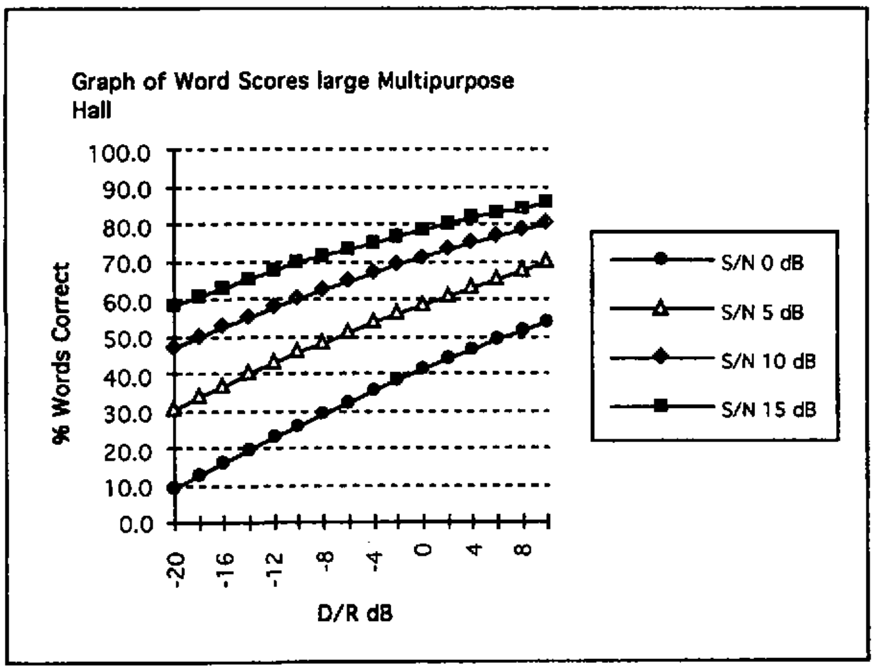

Graph of Word Scores large Multipurpose Hall
% Words Correct
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
-20
-16
-12
-8
-4
0
4
8
D/R dB
S/N 0 dB
S/N 5 dB
S/N 10 dB
S/N 15 dB

Fig 8

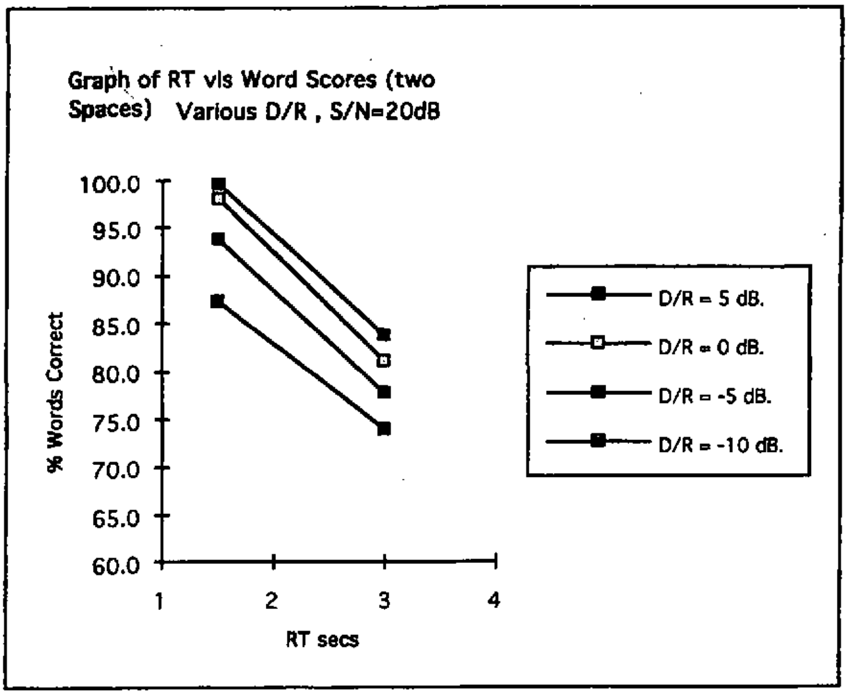

Graph of RT vls Word Scores (two Spaces) Various D/R , S/N=20dB
100.0
95.0
90.0
85.0
80.0
75.0
70.0
65.0
60.0
% Words Correct
1
2
3
4
RT secs
D/R = 5 dB.
D/R = 0 dB.
D/R = -5 dB.
D/R = -10 dB.

Fig 9

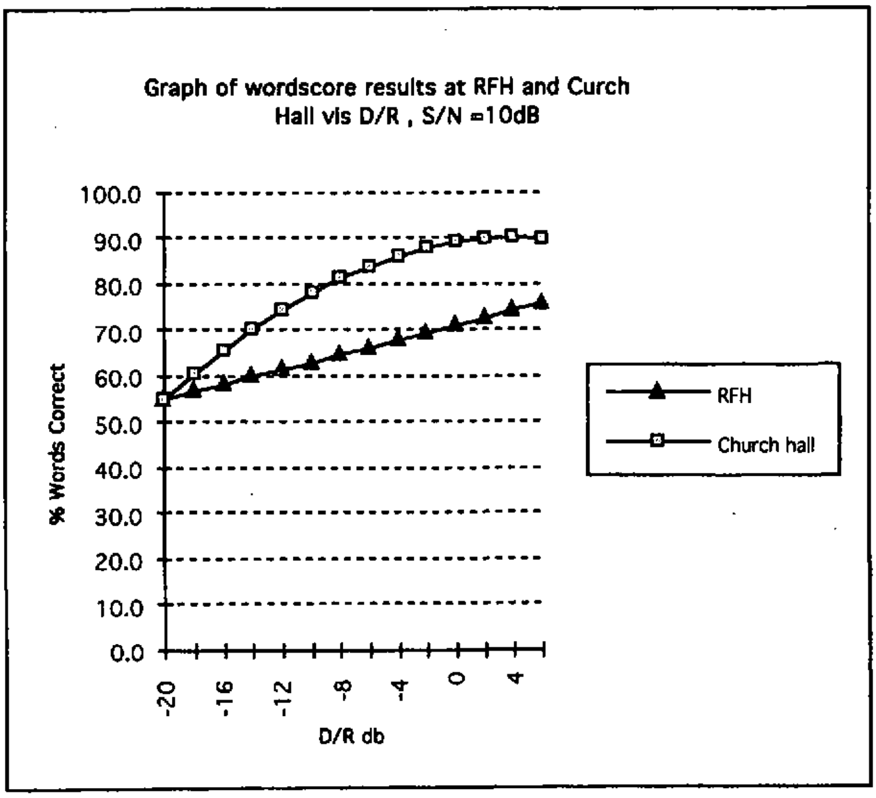
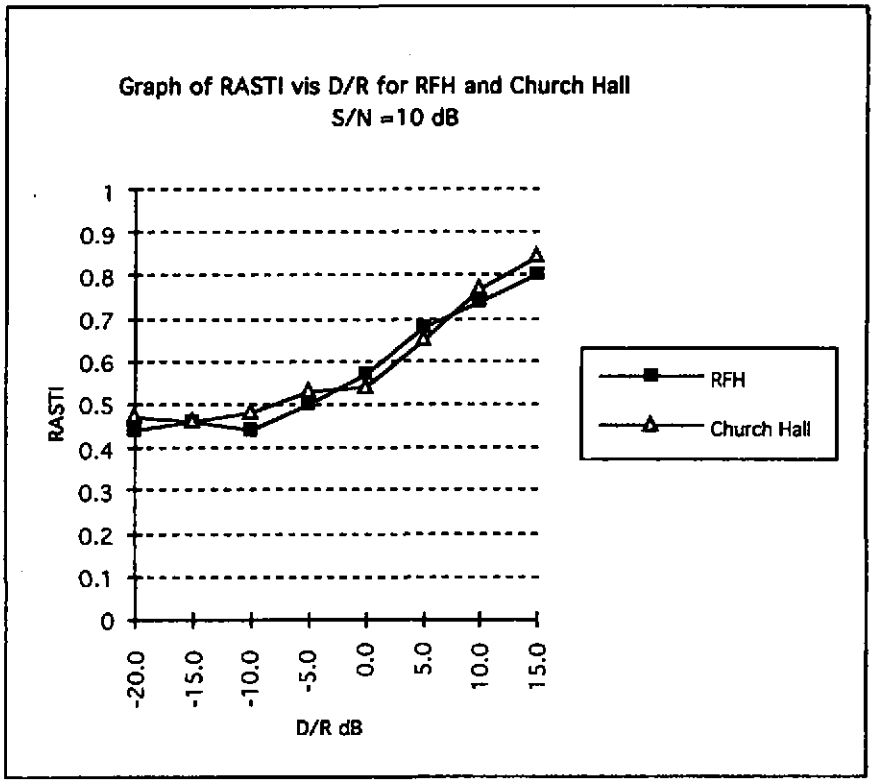
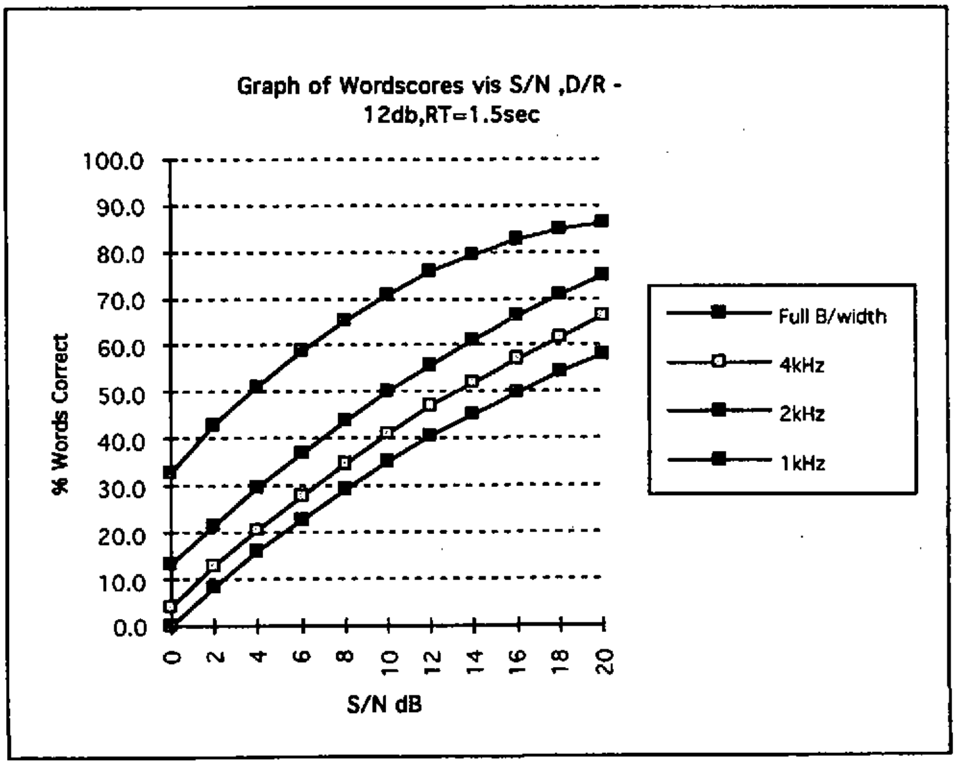

Graph of wordscore results at RFH and Curch Hall vls D/R , S/N =10dB
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
% Words Correct
-20
-16
-12
-8
-4
0
4
D/R db
RFH
Church hall

Fig 10



Graph of RASTI vis D/R for RFH and Church Hall
S/N =10 dB
RASTI
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-20.0
-15.0
-10.0
-5.0
0.0
5.0
10.0
15.0
D/R dB
RFH
Church Hall

Fig 11



Graph of Wordscores vis S/N ,D/R -12db,RT=1.5sec
% Words Correct
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0
2
4
6
8
10
12
14
16
18
20
S/N dB
Full B/width
4kHz
2kHz
1kHz

Fig 12

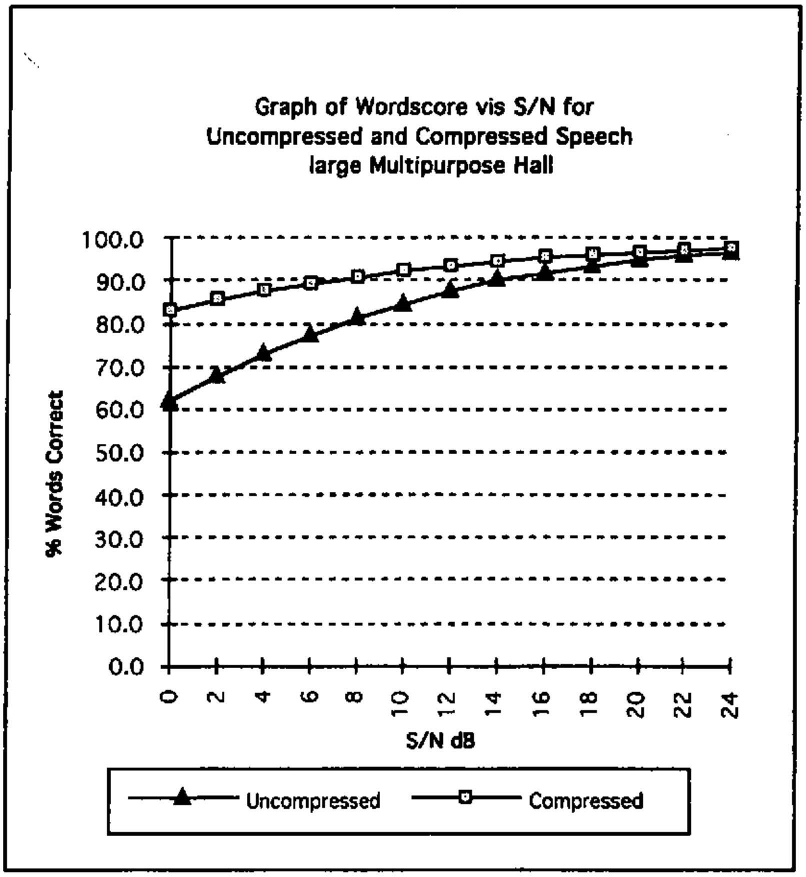

Graph of Wordscore vis S/N for
Uncompressed and Compressed Speech
large Multipurpose Hall
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
% Words Correct
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Uncompressed
Compressed

Fig 13

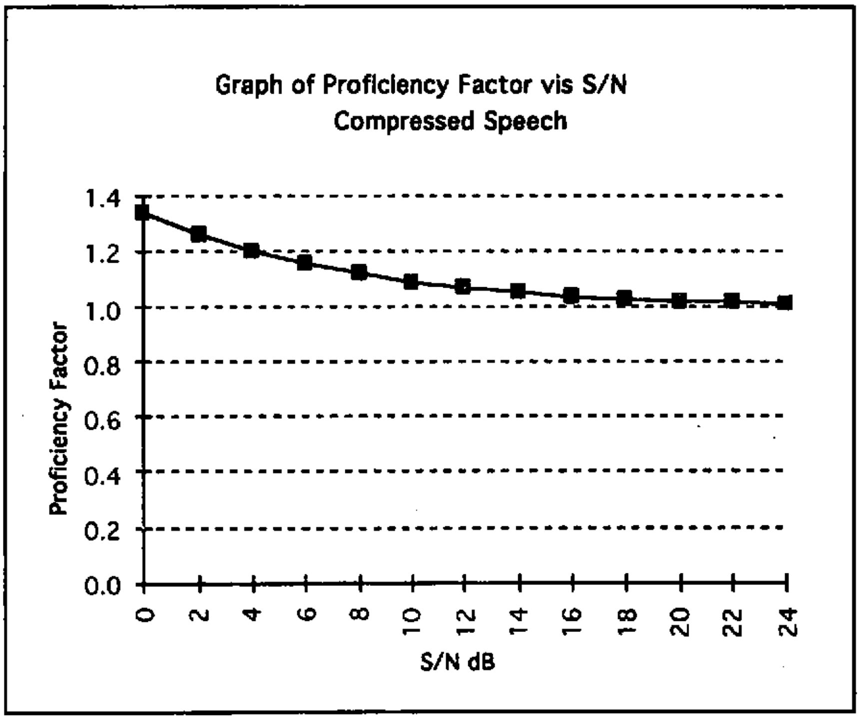

Graph of Proficiency Factor vis S/N
Compressed Speech
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Proficiency Factor
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB

Fig 14

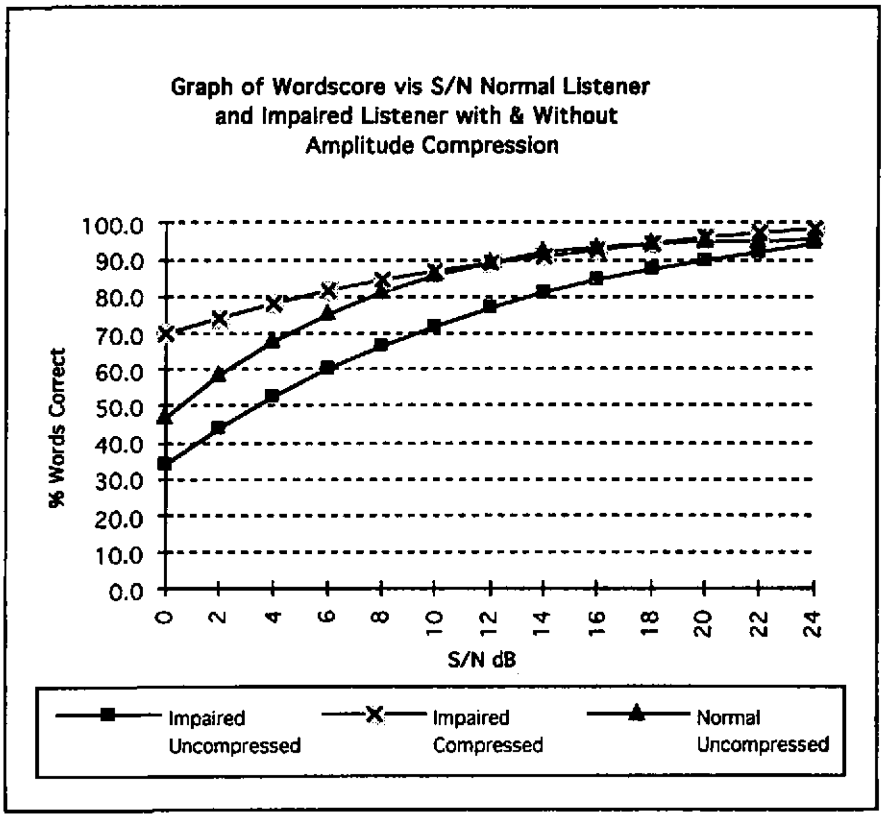

Graph of Wordscore vis S/N Normal Listener and Impaired Listener with & Without Amplitude Compression
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
% Words Correct
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Impaired Uncompressed
Impaired Compressed
Normal Uncompressed

Fig 15

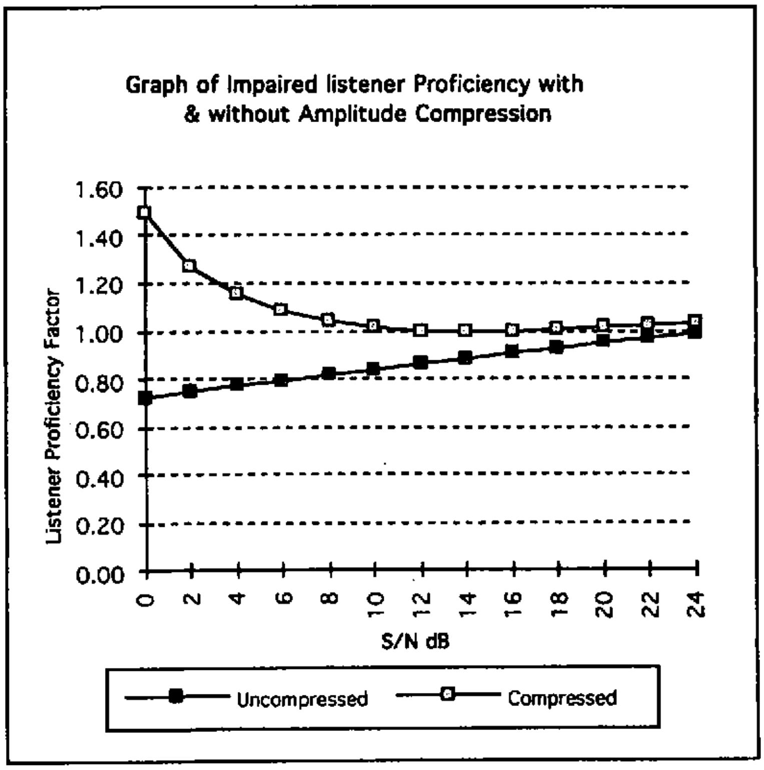

Graph of Impaired listener Proficiency with & without Amplitude Compression
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Listener Proficiency Factor
0
2
4
6
8
10
12
14
16
18
20
22
24
S/N dB
Uncompressed
Compressed

Fig 16