Proceedings of the Institute of Acoustics

PRELIMINARY FINDINGS OF RESEARCH INTO THE EFFECT OF AMPLITUDE COMPRESSION ON SPEECH INTELLIGIBILITY IN THE PRESENCE OF NOISE AND REVERBERATION

Peter Barnett

AMS Acoustics, 21 Queen Annes Place, Bush Hill Park, EN1 2QB.

INTRODUCTION

This Paper presents the progress thus far in a research project to quantify the improvement in speech intelligibility following the application of amplitude compression.

This research project which is partly funded by the Dti 'Smart' competition award scheme involves the recording of both compressed and uncompressed word lists in a variety of spaces.

The spaces used thus far with their outline acoustic and architectural statistics are shown in the table below:

Space	Volume (m^3)	RT_{500Hz} (secs.)	Source
Church Hall	815	1.3	Omni
Concert Hall	21950	1.4	Omni
Football Stadium	N/A	3.0	PA

We are in the process of carrying out measurements in other spaces.

METHOD

In principle the idea was to binaurally record both compressed and uncompressed word lists at various known locations in each space and present, following post-processing, the word lists to a panel of listeners. The word lists were output to the space through either the installed PA system or through a Bruel & Kjaer omni-source. The results of the compressed and uncompressed lists were then compared.

Source material - two talkers, one male and one female, were trained in accordance with ISO TR4870 in the delivery of word lists. Each talker then after training recorded twenty lists of 50 No. words each in alphabetical order in BRE anechoic chamber. This ensured that the recordings were noise-free and without any reverberant components.

Both PB words and CVC lists were used.

The target word was delivered in a carrier sentence as: 'Please Write BAT Now or Please Write TIB Now.

The phrase preceding the target word allows excitation of the reverberant field and prepares the listener for the word that is coming, and the following word, generally 'Now" assists the talker in correctly delivering and voicing the target word.

The recordings were made with a high quality studio microphone to a professional DAT recorder.

These recordings were then digitally read into a digital editing system, to be normalised.

Normalisation is an important process to ensure the consistency of the source material. The normalisation process is as follows:

Each phrase 'Please Write BAT Now' is level checked by comparing the 'Please Write' portion only, each target word has its level adjusted in relation with the 'Please Write', this is important to preserve the true spoken level of the target word.

Each phrase is then stripped to its component parts 'Please Write' space 'BAT' - space 'Now'.

A good sample of 'Please Write' and 'Now' was then reassembled to each of the target words with a precise interstitial space of digital silence.

The entire list was then reassembled at a precise repetition rate using the same length of digital silence between each phrase.

The normalised lists were then consigned to CDR, the transfer being carried out in the digital domain. The first few tracks of the CD contained various signals including pink noise, maximum length sequence and octave band tone bursts to facilitate objective acoustical tests.

COMPRESSION

Each word list was duplicated digitally and the duplicate list compressed with the required compression ratio and threshold, the act of compression introduced an overall gain change which was subject to restoration.

Site Recordings

The site recordings were made in varying known positions in each space using a Bruel & Kjaer Head and Torso Simulator connected to DAT medium.

Several lists of both CVC and PB Words were recorded at each position. Additionally a binaural recording of pink noise was made in the space at a position removed from the source that placed it deep into the reverberant field. In this way the noise 'took on' the acoustic attributes of the space. At each position a full set of acoustic tests were carried out to determine RT, D/R and STI.

Post Processing

The post-processing involved the addition of noise at each measurement position at various, generally three, signal-to-noise ratios. At the same time the words were shuffled in readiness for the listening panel.

The actual process of post-processing is as follows:

The site DAT recordings were read back into the digital editor again remaining in the digital domain throughout its transfer. The 'Please Write' and the 'Now' were stripped off each phrase. Each word was assigned a file and the files shuffled using a random sequence generator. Generally four shuffles, one of each list both compressed and uncompressed, were made. The random lists were then re-assembled using the same procedure as for the source lists but with the exception that the word 'Now' is omitted since it had no further part to play and it only served to increase the listening panel time and possibly introduce additional fatigue and boredom. On this matter also, the space between phrases was reduced but still giving sufficient time to write the word down. This feature of our word scoring programme borne out of experience.

Finally, prior to being consigned to CDR for listening panel tests, each shuffle was subjected to the addition of noise using the 'reverberant' noise recorded binaurally in the space. Generally the following signal-to-noise ratios were applied 0dB, +10dB, +20dB and no noise. It should be understood that the noise was never subjected to compression.

In the addition of the noise, great care was taken to ensure that the signal-to-noise ratios, determined from the 'Please Write' were identical for both the compressed and uncompressed lists.

Listening Tests

The listening tests were conducted using a panel of at least ten persons with hearing judged to be in the range 'normal'. In fact, we used two panels, the other had hearing which was judged to be in a broad sense 'impaired'.

The normal hearing group comprised mainly person below 30 but with one exception of a female who is a practised listener giving consistent results whose age is 50. The hearing impaired group were all over the age of 50 and mostly over the age of 60. A detailed analysis of this group results is the subject of a further Paper.

All listeners had undergone training in accordance with ISO TR4870 and most had previous experience of word score procedures. The listeners were not cognisant of the underlying reason for the tests nor were they aware that compression was involved.

The word lists were presented to the listeners from the CDR over high quality Sennheiser headphones in a quiet environment. Only three lists of 50 No. words each were delivered to the listeners at a single sitting. Between each sitting the listeners were able to relax in a separate common room. In all, a day or listening session comprised 6-8 sittings.

Clearly some tests were more difficult than others and included those that were further from the source with low signal-to-noise ratios.

We categorised the lists as Difficult, Moderate and Easy. We had found from previous experience that a listening sitting sequence DDD, led to erroneous results in the penultimate and final list of the sitting brought about by either fatigue, despair, despondency or irritation.

Accordingly we arranged our sittings to include not more than one D and not more than two M's.

We have some scant evidence that the sequence EMD gives a better result for the D than DME. Accordingly the D test was generally arranged for consistency sake either in the centre or at the end.

We also interspersed control lists of words with varying signal-to-noise ratios for which we know the answer (they had been previously used). When listeners failed to return the expected results on the control tests, the results for that listener for the sitting prior to and after the control tests were placed on the suspect list and in fact generally repeated. In some cases the day results were repeated. In several cases we had to 'let the listener go'. Discrepancies in the control tests were generally found to be attributable to a poor state of mind or temporary distraction.

All listeners were given a hearing test comprising octave band 'staircases' to assess frequency response. We also gave listeners periodic unannounced tests to ensure that they had not attended discos or other high level activities the night before.

Those parties who had to travel by public transport to the tests were issued with ear plugs. All listeners were cognisant of the importance of adverse effect of noise on their ability to carry out the tests and knew that control tests were used but did not know which was which. We would describe our listeners as highly motivated.

RESULTS

The series of experiments has involved some 150 lists scored by up to 18 No. persons. The data obtained shows the difference between compressed and uncompressed words for each position and for each signal-to-noise ratio. A presentation of this data is beyond the scope of this Paper.

To illustrate the effects, I have presented only the data for the female talker for PB Word Tests.

Furthermore, in reporting individual spaces I have only used those obtained in the church.

Figs. 1 to 3 show the results obtained for each position.

To clarify the results, they are represented in fig. 4 as differences between compressed and uncompressed speech.

It can be seen that although there is some understandable uncertainty at low signal-to-noise ratios, a general improvement trend can be seen.

The other spaces show a similar trend as shown in figs. 5 & 6.

It can be seen in general that the improvement (difference uncompressed/compress) increases with reducing S/N and reducing D/R ratio. Of course at low values of S/N and D/R the improvement also reduces since under extreme conditions the words cannot be correctly heard even if compressed.

Finally, it seemed logical that since the improvement is dependent on the absolute magnitude of the returned score, it might be of use if all results irrespective of space or acoustic conditions be compared and graphically presented.

Fig. 7 shows this data.

CONCLUSIONS

The results obtained so far show a worthwhile improvement in correctly scored words following the application of amplitude compression.

This improvement would represent a significant mitigation to situations where either noise or poor acoustics or both was an attendant problem.

The scope of this Paper has not allowed the presentation of all of the data obtained. I can report that similar and sometimes greater improvements have been observed from our hearing-impaired jury tests.

Our Dti Smart Research Project is due for completion by the end of the year when we would expect to be able to publish all of our findings.

Graph of Word Score Results - Small Church at Position 1
Direct-to-Reverberant Ratio =+5dB

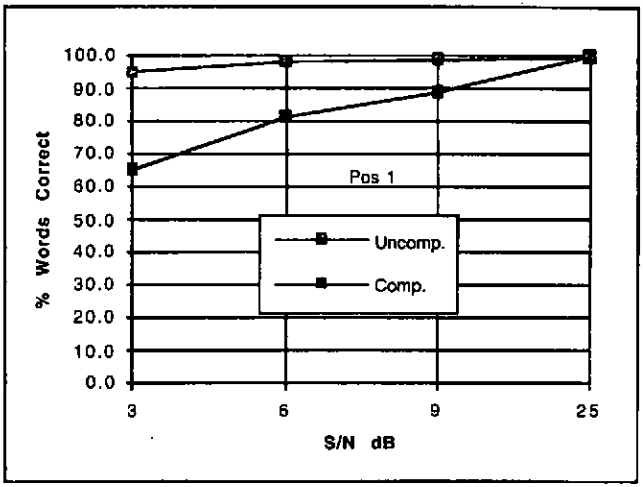

Fig. 1

Graph of Word Score Results - Small Church at Position 2
Direct-to-Reverberant Ratio =0dB

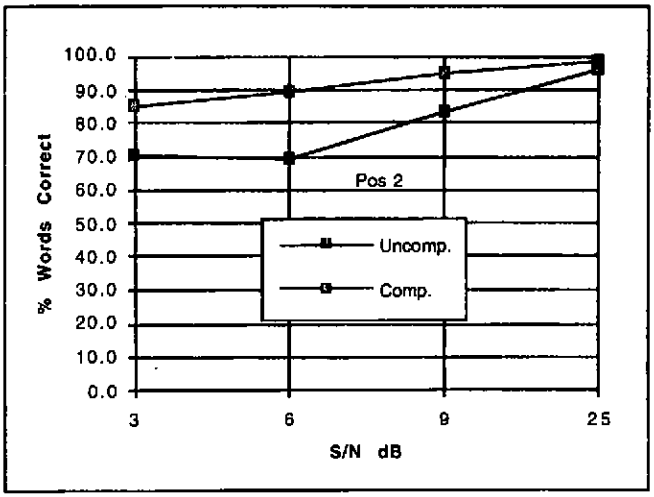

Fig. 2

Graph of Word Score Results - Small Church at Position 3
Direct-to-Reverberant Ratio ≈-10dB

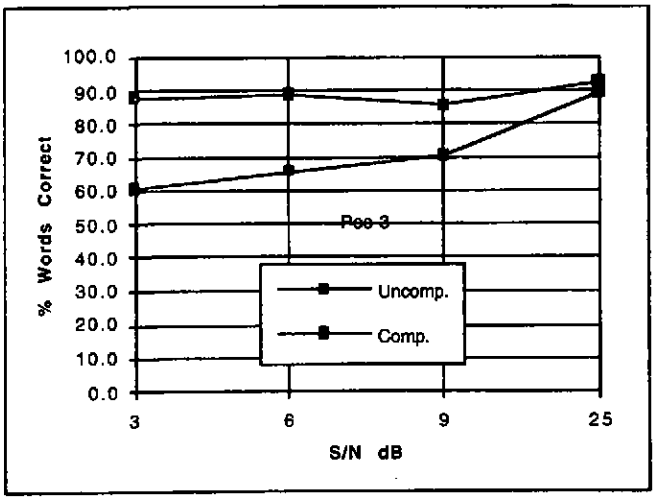

Fig. 3

Graph of Word Score Improvement
(Compressed Score less Uncompressed Score) at Small Church

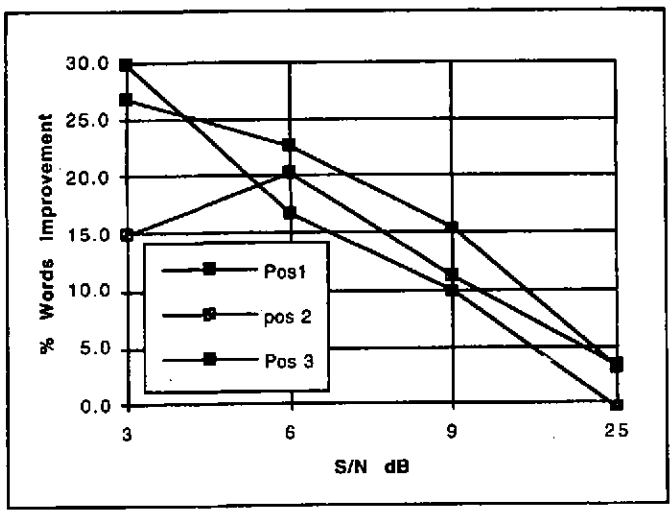

Fig. 4

Graph of Word Score Improvement
(Compressed Score less Uncompressed Score)
at Large Concert Hall

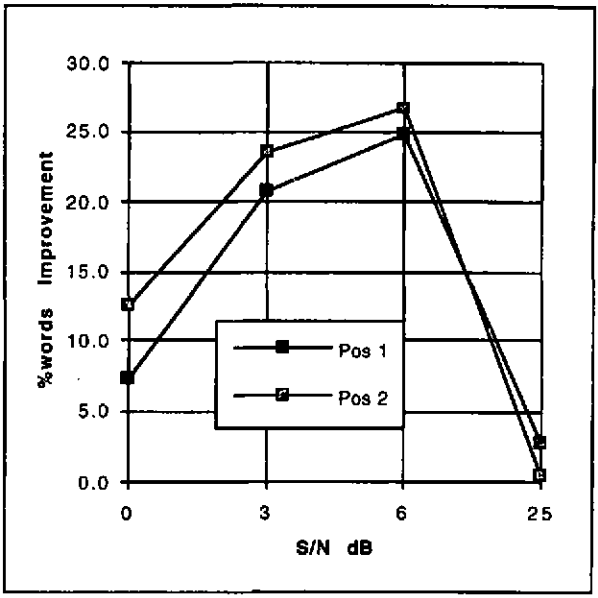

Fig. 5

Graph of Word Score Improvement
(Compressed Score less Uncompressed Score)
at Football Ground

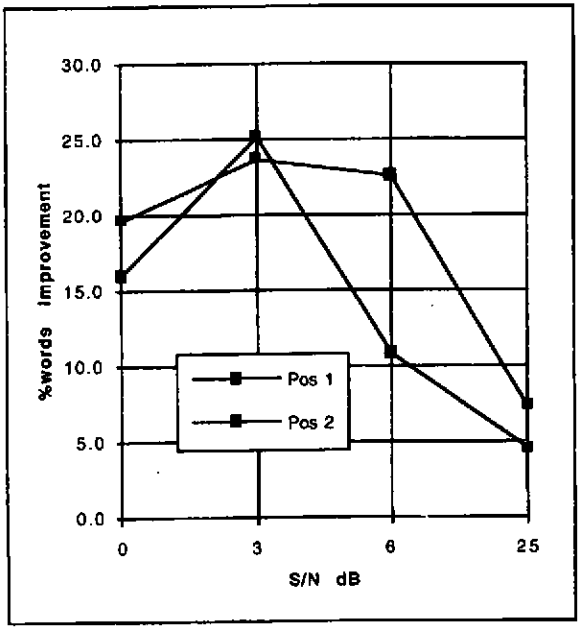

Fig. 6

Graph of % Words Correctly Understood for Both Compressed and uncompressed Signal (All spaces)

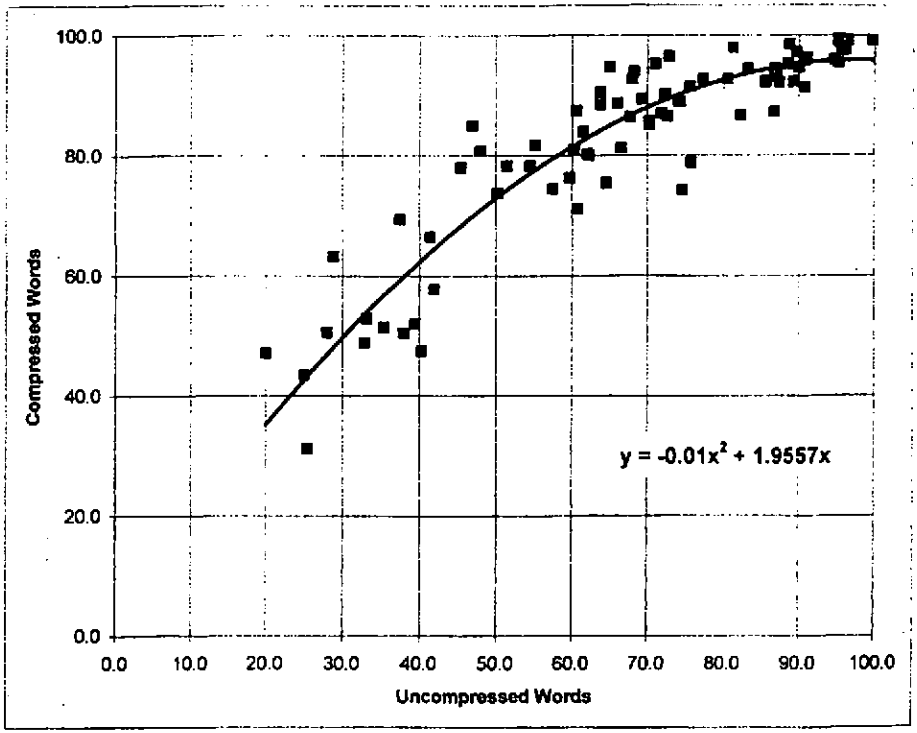

Fig. 7

Graph of Impovement in % Words Correcting Understand Following the Application of Amplitude Compression (All Spaces)

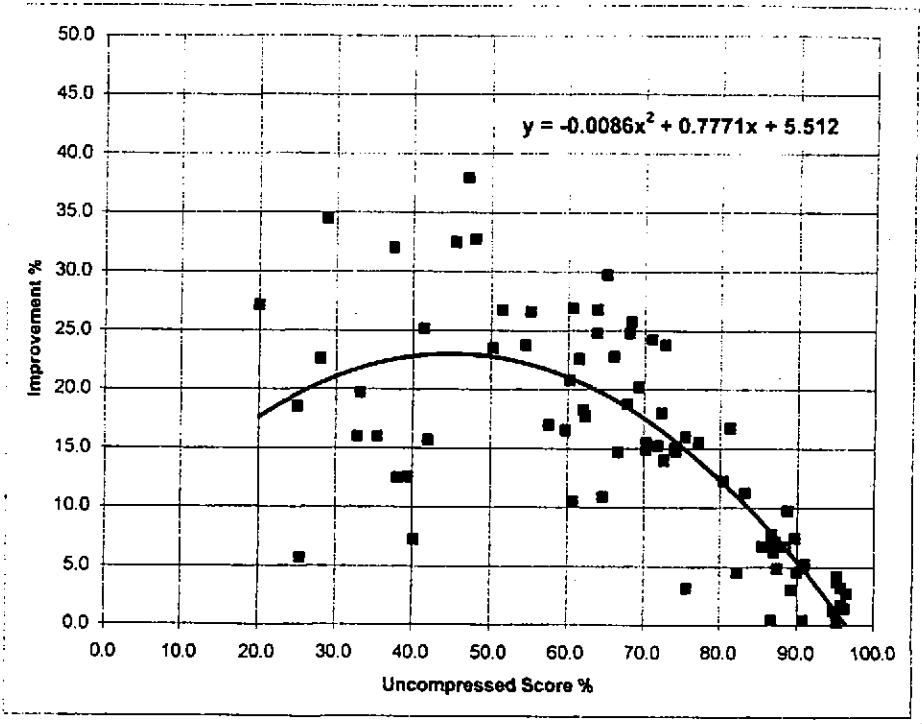

Fig. 8