Proceedings of the Institute of Acoustics

STUDY OF WORD SCORE TEST RESULTS TO DETERMINE THE ROBUST COMPONENTS OF SPEECH SUBJECT TO NOISE AND REVERBERATION

Peter Barnett

AMS Acoustics, Rayleigh House, 21 Queen Annes Place, Enfield, EN1 2QB.

INTRODUCTION

This Paper is concerned with an analysis of Word Score data which was obtained from part of our ongoing experiments and measurements. It was intended to present data obtained from a number of spaces, however the analysis took rather more time than expected and as such the data presented is for experiments carried out in a small church hall (St. Pauls) and a large concert hall (Royal Festival Hall).

The object of the analysis was to determine which components of our speech are robust and which are fragile in the presence of both noise and reverberation. It should be emphasised that this analysis was not the object of the measurements which were generally carried out to determine and scale speech intelligibility.

The data under analysis comprises CVC nonsense scores determined in each space with the addition of noise. In each case the source was a Bruel & Kjaer omni-directional loudspeaker and the word lists were contaminated with noise in the following signal-to-noise ratios >25dB, 6dB, 3dB and 0dB.

The following tables provide the relevant acoustical data for each space.

Acoustic Data for St. Pauls

Parameter	Position No.		
	1	2	3
RT (sec.) 500Hz/2kHz	1.3/1.6	1.2/1.4	1.3/1.5
D/R (dB) 500Hz/2kHz	-3.3/-4.1	-8.6/-9.5	-14.9/-15.6
RASTI	0.54	0.52	0.46

Acoustic Data for The Royal Festival Hall

Parameter	Position No.	
	1	2
RT (sec.) 500Hz/2kHz	1.5/1.5	1.4/1.6
D/R (dB) 500Hz/2kHz	-3.8/-4.1	-12.2/-12.8
RASTI	0.48	0.48

WORD SCORE TEST RESULTS

The methods and procedures used in the preparation and determination of the word tests is given in 'Preliminary Findings of Research into the Effect of Amplitude Compression on Speech Intelligibility in the Presence of Noise and Reverberation'.

ANALYSIS OF DATA

Each word for each word list was analysed to determine which component of the word e.g. start consonant, vowel or end consonant was misheard by the listener.

In addition each word was split into its phonetic components e.g. vowel, (long and short) diphthongs, transitionals, semi-vowels, fricative consonants (voiced, unvoiced) and stop consonants (voiced, unvoiced).

The following abbreviations were used:

Component	Abbreviation
Vowel - Long	VOL
Vowel - Short	VOS
Diphthong	DIP
Semi-Vowel	SVO
Fricative Consonant Vowel	FCV
Fricative Consonant Unvoiced	FCU
Stop Consonant Voiced	SCV
Stop Consonant Unvoiced	SCU
Transitional	TRA

RESULTS

The tables overleaf show the results of the analysis in regard of word start, end and vowel.

Percentage Start, End and Vowel Misheard for St. Pauls

S/N (dB)	Position	Word		
		Start %	End %	Vowel %
25	1	1	4	1
	2	1	2	2
	3	11	6	3
6	1	6	7	1
	2	6	6	3
	3	23	17	5
3	1	8	12	1
	2	10	9	4
	3	29	23	4
0	1	13	14	2
	2	13	13	4
	3	42	35	5

Percentage Start, End and Vowel Misheard for Royal Festival Hall

S/N (dB)	Position	Word		
		Start %	End %	Vowel %
25	1	7	7	2
	2	12	7	1
6	1	37	30	3
	2	49	43	5
3	1	48	35	4
	2	56	48	4
0	1	55	44	6
	2	73	61	7

From the foregoing tables it can be seen that the vowel sounds are far more robust than the consonants.

In regard of the consonant sounds, three effects can be observed.

Firstly, the effect of noise at each position as the S/N reduces there is a steady reduction in the consonants correctly heard. A similar effect may be seen occurring to the vowel sounds.

Secondly, there is a steady reduction in the consonants correctly heard as the direct-to-reverberant ratio decreases. This effect is less prominent for the vowel sounds.

Thirdly, at St. Pauls the emphasis changes from the end of the word misheard to the start. This we believe results from the masking effect of the carrier sentence increasing as the direct-to-reverberant ratio decreases. The fact that this is not seen in the Royal Festival Hall results, we believe, can be explained by the much longer mean free path of the Royal Festival Hall.

Hence we may deduce that the loss of intelligibility is due to

(a) Noise masking

(b) Reverberation masking.

It can be seen that in both cases (S/N and D/R) the mechanism is similar, except that in the presence of reverberation, the effect is more deleterious.

The following tables show the percentage of phonetic components incorrectly heard:

WORD SCORE TEST RESULTS

Percentage of Phonetic Sounds Misheard
at the Start of the Word in St. Pauls

S/N (dB)	Position	Phonetic Component					
		SCV %	SVU %	FCV %	FCU %	TRA %	SV0 %
25	1	0	5	0	0	6	0
	2	2	1	8	2	5	0
	3	4	5	25	3	33	2
6	1	2	3	0	0	35	0
	2	5	3	23	6	7	2
	3	6	21	28	8	50	23
3	1	3	5	3	3	33	1
	2	6	13	25	10	9	4
	3	10	26	30	13	56	27
0	1	9	7	10	8	48	2
	2	14	14	28	13	20	7
	3	15	34	35	20	59	35

Percentage of Phonetic Sounds Misheard
at the Start of the Word in the Royal Festival Hall

S/N (dB)	Position	Phonetic Component					
		SCV %	SVU %	FCV %	FCU %	TRA %	SV0 %
25	1	1	5	13	13	10	11
	2	2	5	0	20	25	9
6	1	9	7	23	18	40	18
	2	9	6	0	38	52	11
3	1	15	11	30	18	50	33
	2	18	9	3	30	50	26
0	1	34	28	43	40	83	53
	2	42	34	15	33	83	38

WORD SCORE TEST RESULTS

Percentage of Phonetic Sounds Misheard
at the End of the Word in St. Pauls

S/N (dB)	Position	Phonetic Component				
		SCV %	SVU %	FCV %	FCU %	SV %
25	1	2	0	1	2	0
	2	3	0	4	2	1
	3	8	5	13	7	1
6	1	6	4	6	6	0
	2	10	8	16	7	1
	3	29	13	47	14	6
3	1	8	7	18	9	3
	2	11	10	29	11	5
	3	31	18	51	19	10
0	1	15	14	28	10	4
	2	22	17	34	13	6
	3	38	25	58	31	14

Percentage of Phonetic Sounds Misheard
at the end of the Word in the Royal Festival Hall

S/N (dB)	Position	Phonetic Component				
		SCV %	SVU %	FCV %	FCU %	SV %
25	1	5	13	9	9	3
	2	2	10	13	17	4
6	1	32	25	61	41	3
	2	42	47	60	56	9
3	1	35	29	76	41	1
	2	52	51	73	69	9
0	1	51	35	72	69	8
	2	67	63	80	70	15

It can be seen that there is a difference between the components at the start and end of the word.

For the start of the word, semi-vowels, voiced stop consonants and unvoiced fricative consonants seem to be preferred.

Transitionals and voiced fricative consonants are to be avoided.

In regard of the end of the words, voiced fricative consonants are to be avoided and semi-vowels or stop consonants are preferred.

It is also possible to deduce that voiced stop consonants and unvoiced fricative consonants are less affected by reverberation. The same, but to a lesser effect is true of unvoiced stop consonants and semi-vowels.

DISCUSSION

This limited analysis does seem to suggest that there is some merit in pursuing the matter and as such, we intend to expand our analysis to include other spaces.

The value in the endeavour is patent. In acoustically difficult situations it might be possible to construct emergency messages using mainly robust speech components.

This limited study has identified certain speech sounds to be avoided and as such this is a start. Our next endeavour is to preferentially select words from the accepted word score test population which should give a bias towards the two extremes. In this way we might be able to quantify this preferential effect. We also intend to examine the difference between CVC nonsense and PB Words to establish recognition and if the recognition has a preference for certain speech sounds.

Unfortunately, subjective testing is a lengthy and costly process and hence it may take a little time.

CONCLUSIONS

We have formed the following conclusions:

1. Vowel sounds are more robust than consonants.

2. At the start of the word
   Semi-vowels, voiced stop consonants and unvoiced fricatives are to be preferred.
   Transitionals and vowel fricatives are to be avoided.

3. At the end of the words
   Semi-vowels are preferred.
   Voiced fricatives are to be avoided.

4. There is merit in developing this strategy.