

# A COMPARISON OF SUBJECTIVE SPEECH INTELLIGIBILITY TESTS IN REVERBERANT ENVIRONMENTS

**K. Kruger**  
Alberta Public Works  
Acoustics Section

**K. Gough**  
University of Alberta  
Dept. of Educational Psychology

**P. Hill**  
Educational Consultants  
for the Sensory Impaired

## INTRODUCTION

Subjective speech intelligibility tests are not often used because of their time consuming nature. However, a lack of accurate objective measurement techniques requires that a subjective approach be used in many critical situations. This is especially true in large reverberant spaces where the primary sound source is a loudspeaker system.

For subjective evaluation of articulation it is necessary to use a test that is reasonably simple to conduct, is sensitive to changes in the acoustical parameters that affect intelligibility and is indicative of intelligibility of running speech occurring under realistic conditions. Other studies<sup>1,2,3</sup> that have investigated the effects of reverberation and signal-to-noise on intelligibility have used a variety of subjective measurement instruments including Fairbanks Rhyme Test, Modified Rhyme Test, Phonetically Balanced Word Tests, CVC Word Tests and spondaic word tests. It is recognized that the various tests lead to different results but the intrinsic relationship between them has not been clearly established for the case where the primary distorting factor is reverberation. It is often assumed that reverberation affects discrimination in a similar manner as signal-to-noise ratio.

This study compared intelligibility scores for three subjective test methods against a wide range of reverberant conditions. The three subjective methods were CID Phonetically Balanced Word Test (PB), Modified Rhyme Test (MRT) and CID Test Sentences. Reverberation times varied from approximately 1 second to 6 seconds. Other aspects, not typically explored in previous studies, included a rating scale of difficulty, a measure of the subjects comprehension of speech, and a phonemic analysis of PB wordlist errors.

## PROCEDURE

The stimuli for each test were recorded in near anechoic conditions using a single male speaker. Reverberation was added to the signal upon presentation to the listeners by routing the pre-recorded signal through a digital multi-effect processor. With the processor it was possible to control the relative intensities of the direct sound, early and reverberant energy and the latter portion of the decay. It was decided to simulate an ideal exponential decay as closely as possible. Figure 1 illustrates the reverberation time versus frequency for the various test conditions.

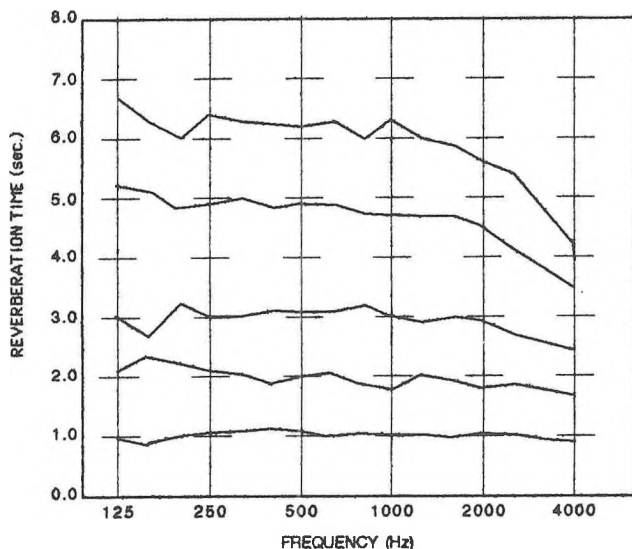


Figure 1 Reverberation time versus frequency for five test conditions.

Five subjects, 22 - 28 years old, were screened for normal hearing and speech discrimination. All pure tone threshold scores fell within a range of 0 dB to 20 dB (ISO). Speech discrimination scores using PB-50 word lists were 100% across all subjects.

The articulation tests were presented through headphones. Subjects first listened to samples of each speech stimulus and were given the opportunity to adjust the loudness to their Most Comfortable Level (MCL).

During the administration of PB words, subjects were required to state aloud the perceived stimulus words which were recorded as correct or incorrect by the test administrator. Subjects responded to the MRT stimuli by circling the perceived word on a response sheet that listed the target word with 5 distractors per stimulus. Each subject responded to the Sentence test by repeating the sentence which was transcribed by the administrator as well as being recorded on audio-tape for future reference. Each test (PB words, MRT, and Sentence test) was administered at each of five reverberant conditions. This generated fifteen different listening tasks which were presented to the subjects in a randomly assigned order.

At the completion of the tests, subjects were asked to judge the level of difficulty they experienced in performing the task. In addition, the subjects comprehension of speech was determined through their responses to questions on short anecdotes presented at each reverberant condition.

## RESULTS AND DISCUSSION

Figure 2 shows the mean discrimination scores for each of the three articulation tests. The indicated reverberation times represent the averages over the 500 Hz to 2000 Hz one-third octave bands with the characteristics shown on Figure 1. The PB and Sentence tests indicate the greatest sensitivity to increases in reverberation time. Of the two, the PB test was also found to have the least variability in scores at each condition. The MRT showed minimal sensitivity. Over the extreme range of reverberant conditions, the modified rhyme scores varied by only 13%. The PB and Sentence tests in contrast varied by 49% and 38% respectively.

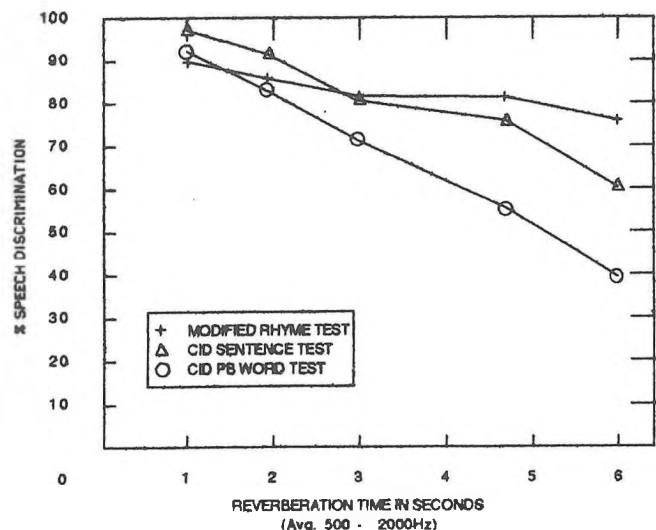


Figure 2 Mean speech discrimination scores versus reverberation times for various articulation tests.

Figure 3 shows the relationship between the same tests but under various signal-to-noise conditions<sup>4</sup>. It appears that there are both similarities and differences between the articulation tests for contaminating factors of background noise and reverberation. In both cases the PB and Sentence tests are the most sensitive to changes. Also the Sentence test scores are consistently higher for all conditions. The marked difference relates to the curves for the MRT. In the case of noise, the MRT displays reasonable sensitivity over the S/N conditions and the absolute value of the discrimination scores tends to be less than the Sentence test. Under reverberation, however, the MRT differentiates very little over a wide range of conditions. It is hypothesized that the MRT exhibits this characteristic because of two reasons. Firstly, the MRT does not have a carrier phase so that there is no reverberant masking from previously spoken words to interfere with discrimination. Secondly, 50% of the test word groups requires that the listener correctly identify the initial phonemes. Since there is no self masking taking place during the initial word pulse there are no contaminating factors to inhibit correct identification.

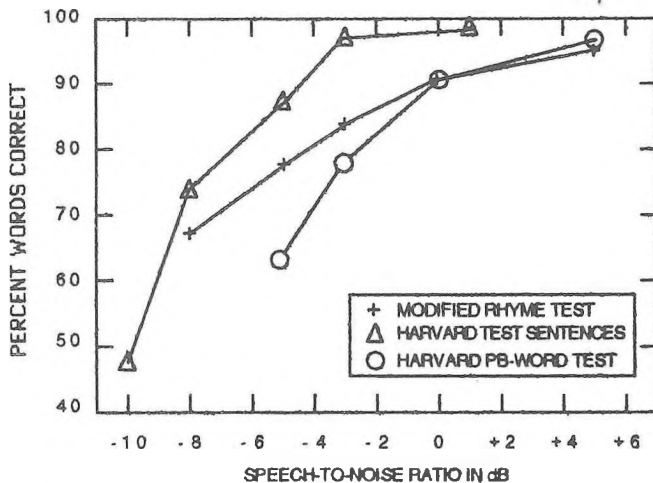


Figure 3 Discrimination score versus speech-to-noise ratio for various articulation tests. <sup>4</sup>

PB word tests appear to be sensitive to both noise and reverberation conditions, particularly when the listeners are responding to 50-word lists. Typically the scoring protocol simply involves determining the percentage of words correctly identified. This scoring system does not indicate, however, the number and type of phonemic errors. This type of analysis may provide more specific information on the effects of reverberation and background noise. From the results of the PB tests, the specific phonemic elements of the incorrect words were categorized in terms of phoneme class for the consonants and vowels under each of the reverberant conditions. Figure 4 presents the overall vowel and consonant errors.

The difficulty test scale rated from 1 (no difficulty) to 5 (extreme difficulty) showed, not surprisingly, that difficulty increased for all tests as reverberation increased. The judgments of least difficulty were attributed to the MRT where the highest difficulty level was 3 compared to 5 for the PB and MRT at the extreme reverberation time. The anecdotal test of comprehension indicated the highest correlation with the PB test scores.

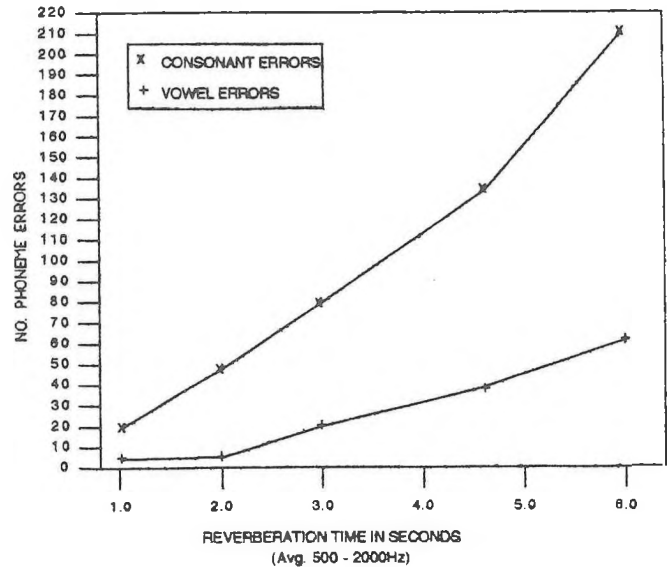


Figure 4 Phonemic errors from PB word test.

## CONCLUSION

Of the three subjective measurement instruments considered in this pilot study, the PB word test was the most responsive to changes in reverberation. This finding agrees with studies that found similar results under signal to noise conditions. The Sentence test compares favourably in terms of sensitivity with the PB word tests and closely resembles the trends found under noise conditions. The MRT was determined to be a poor indicator on the effects of reverberation on speech intelligibility. The study also provided information on aspects of intelligibility not usually considered in studies of this type. This included listeners' subjective impressions of level of test difficulty, auditory comprehension under reverberant conditions and a phonemic analysis of subjects' responses to PB tests. Further support for the PB test as the best of the three measures of intelligibility was indicated by these results.

## REFERENCES

- 1 Nabelek, A.K. and Robinson, P.K. Monaural and Binaural Speech Perception in Reverberation for Listeners of Various Ages. *J. Acoust. Soc. Am.* 71(5), 1242 - 1248 (1982)
- 2 Houtgast, T. and Steeneken, H.J.M. A Multi Language Evaluation of the RASTI - Method for Estimating Speech Intelligibility in Auditoria, *Acustica* 54(4) 185 - 199 (1984)
- 3 Bradley, J.S. Uniform Derivation of Optimum Conditions for Speech In Rooms, National Research Council Canada, BRN 239
- 4 Williams, C.E. and Heckler, M.H.L. Relation Between Intelligibility Scores for Four Test Methods and Three Types of Distortion, *J. Acoust. Soc. Am.* 44(4), 1002 -1006 (1968)