PAN ARCHIVES OF ACOUSTICS
Vol. **37**, No. 2, pp. 199–203 (2012)

DOI: 10.2478/v10168-012-0026-3

Effects of Speech Intensity on the Callsign Acquisition Test (CAT) and Modified Rhyme Test (MRT) Presented in Noise

Misty BLUE-TERRY[(1)], Maranda McBRIDE[(2)], Tomasz LETOWSKI[(3)]

[(1)] *Department of Industrial and Systems Engineering, North Carolina Agricultural & Technical State University*
1601 E. Market Street, Greensboro, North Carolina 27411, USA
Current affiliation: *The Cooperative Extension Program, Coltrane Hall*
P.O. Box 21928, Greensboro, North Carolina 27420-1928, USA

[(2)] *Department of Management, North Carolina Agricultural & Technical State University*
1601 E. Market Street, Greensboro, North Carolina 27411, USA; e-mail: mcbride@ncat.edu

[(3)] *U.S. Army Research Laboratory, Human Research & Engineering Directorate*
AMSRD-ARL-HR/B-520, Aberdeen Proving Ground, Maryland 21005, USA

(*received October 24, 2011; accepted April 2, 2012*)

This study sought to evaluate the effect of speech intensity on performance of the Callsign Acquisition Test (CAT) and Modified Rhyme Test (MRT) presented in noise. Fourteen normally hearing listeners performed both tests in 65 dB A white background noise. Speech intensity varied while background noise remained constant to form speech-to-noise ratios (SNRs) of −18, −15, −12, −9, and −6 dB. Results showed that CAT recognition scores were significantly higher than MRT scores at the same SNRs; however, the scores from both tests were highly correlated and their relationship for the SNRs tested can be expressed by a simple linear function. The concept of CAT can be easily ported to other languages for testing speech communication under adverse listening conditions.

Keywords: speech intelligibility, speech intensity, speech-to-noise ratio.

1. Introduction

Speech intelligibility (SI) is defined as the percentage of speech units (i.e., phonemes, syllables, words, phrases, or sentences) that may be correctly identified by a listener (Letowski *et al.*, 2001). Several different SI tests are currently utilized in both research and practice. One of the more recent English language SI tests is the Callsign Acquisition Test (CAT) developed by the United States Army Research Laboratory. The CAT has similar general applications as the Modified Rhyme Test (MRT) (Fairbanks, 1958; House *et al.*, 1965), which is widely used for assessing SI in easy to moderate speech communication conditions. The primary goal of the CAT is to predict SI of military communications in difficult listening environments characterized by poor signal-to-noise ratios.

The current version of CAT has been used in several studies and evaluated by multiple researchers (Blue *et al.*, 2004; 2010; Rao, Letowski, 2006); however, since it is a relatively new instrument, it is still lacking full validation and standardization. The standardization process of any new SI test involves, among other things, determining test validity and sensitivity; evaluating the effects of noise, talker's voice, and listening environment; and comparing its scores to scores obtained with existing SI tests. Various technical and procedural factors that affect the scores obtained with any SI test material include speech intensity level and speech-to-noise ratio (SNR). The objective of the present study was to measure the effectiveness of the CAT and compare it to the MRT across various SNRs in the presence of a 65 dB A white noise.

2. Methodology

2.1. Participants

A total of 14 normally hearing listeners participated in the study. Normal hearing was defined as pure-tone hearing thresholds at or below 20 dB HL at audiometric octave frequencies from 250 through 8000 Hz. The group was comprised of 8 male and 6 female listeners between the ages of 18 and 25 years.

2.2. Instrumentation

The study was conducted in an Industrial Acoustic Company (IAC) 143M audiometric booth. Instrumentation for the research included (1) a Dell IBM PC/586 computer with a CD ROM drive, (2) two Hewlett-Packard HP-350D step attenuators, (3) a Crown D-75 power amplifier, (4) a CD ROM with test materials and in-house CAT and MRT software for speech signal delivery and data collection, and (5) a pair of AKG K-1000 earphones. A KEMAR (Knowles Electronic Manikin for Acoustic Research) simulator with a Zwislocki coupler (ANSI S3.25) was used to measure sound pressure levels generated at the ear of the listener.

The test materials were installed on the Dell computer. Both the speech signals and noise were played through a multi-channel sound card (Turtle Beach – Santa Cruz). The speech signals were played through one channel and the noise signal was played through the second channel. The speech and noise levels were controlled by two independent HP-350D step attenuators. Once both sound pressure level files were adjusted to the proper levels, they were played through the earphones.

2.3. Test materials

The Modified Rhyme Test (MRT) (Fairbanks, 1958; House *et al.*, 1965) is the most frequently used SI test for evaluating transmission capabilities of acoustic and audio systems. The test uses a battery of 50 sets of 6 one-syllable rhyming or similar sounding words to test initial and final consonant recognition. During the test, one of the words from the list is presented to the listener verbally and the listener is required to indicate which one of the six words in the list was presented.

The Callsign Acquisition Test (CAT) was developed by the United States Army Research Laboratory in response to criticisms that widely used SI testing materials are not effective in certain contexts, particularly military environments, which are noisy and characterized by limited vocabulary communications. Several authors have reported that the use of military personnel in SI studies requires military-specific test material to generate reliable scores (Rao, Letowski, 2006; Howes, 1957). The CAT combines two-syllable words based on the phonetic alphabet with one syllable numeric digits to form a total of 126 three-syllable alphanumeric calling phrases (callsigns). They constitute a family of test items that is familiar to both military personnel and civilians, making it useful both inside and outside of military environments.

Both the MRT and CAT recordings used in the study were made at the U.S. Army Research Laboratory by the same native English male talker speaking with a Midwestern accent. The listeners were familiarized with both tests' materials prior to the study to avoid learning curve effects and to make both tests equally familiar to the listeners since research has shown that familiarity with the test material results in higher and more stable SI test scores (Howes, 1957; Morton, 1969; Schultz, 1964).

2.4. Noise and speech levels

White noise presented at a constant 65 dB A level was used in the study as background noise. The level was selected to be close to the normal conversational level of speech so the naturally spoken speech materials could be used to produce small SNRs. The MRT and CAT test items were presented at five different intensities – 47, 50, 53, 56, and 59 dB A – resulting in SNRs of −18, −15, −12, −9, and −6 dB. Speech levels were determined by averaging dB A levels measured separately for each word.

2.5. Procedure

Each listener was seated inside the acoustically treated booth facing a monitor and keyboard and wearing earphones. Prior to data collection, each participant read written instructions of their tasks, was familiarized with the test material, and given 10 practice trials for each speech test.

The listener's task was to listen to the series of CAT or MRT items and use the appropriate computer screen interface (shown in Fig. 1) and keyboard to record their responses of what they heard. They were instructed to

Fig. 1. Screen captures of interfaces for CAT (top) and MRT (bottom) software.

identify the words that they heard using the keyboard. For example, if the listener heard "Zulu Two" from the CAT, the correct response would be "Z2; Enter". Pressing the "Enter" key would store their response as well as start the next trial. For the MRT, if the listener heard the word "din" from the list as it appears in Fig. 1, the correct response was "5; Enter". If they were unsure of what they heard, they were instructed to make their best guess.

3. Results

Table 1 shows the means (M), standard deviations (SD), and coefficients of variation (V) for each experimental condition. The performance-intensity (PI) functions describing the relationship between the speech intelligibility score and SNR for both the MRT and CAT are shown in Fig. 2. To determine statistical significance, all percentage scores were transformed into rau units (Studebaker, 1985) in order to eliminate the potential of ceiling effects associated with the SI scale. An alpha level of 0.05 was used to determine significance for all statistical tests. A two-factor ANOVA shows that the type of test had a significant effect on the SI performance [$F(1{,}130) = 163.00$, $p < 0.001$] as did the SNR [$F(4{,}130) = 43.95$, $p < 0.001$]; however, there was no significant interaction between the two [$F(4{,}130) = 1.08$, $p = 0.367$].

Table 1. Mean (M), standard deviation (SD), and coefficient of variation (V) for CAT and MRT.

SNR [dB]		−18	−15	−12	−9	−6
CAT	M	65.57	77.38	86.64	97.07	98.93
	SD	17.03	18.44	14.09	3.17	2.3
	V	26.42	29.02	24.97	7.78	4.88
MRT	M	35.79	48.71	61.29	68.86	78.43
	SD	17.76	18.52	16.49	15.42	13.93
	V	48.70	36.09	26.38	23.13	20.62

NOTE: Coefficient of variation (V) has been calculated using *rau* scores.

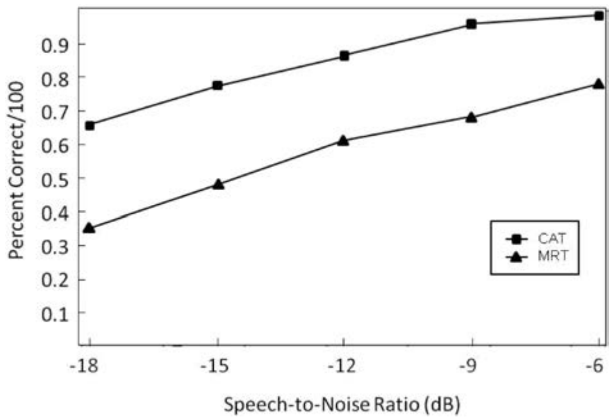

Fig. 2. Perfomance-intensity functions for CAT and MRT.

A correlation analysis was performed to evaluate the relationship between the CAT and MRT performance scores in the tested range of SNRs. The Pearson's correlation coefficient shows that the two tests have a high positive relationship [$r(12) = 0.84$, $p < 0.001$], which validates the parallel shift in the PI functions shown in Fig. 2.

4. Discussion

Based on Fig. 2 and the correlation analysis results, both PI functions have similar shapes and slopes in the tested range of SNRs. The SI performance for the CAT increases by about 3–5%/dB SNR from SNRs −18 to −12 dB before beginning to plateau. Simarly, the MRT increases by about 2–4%/dB SNR throuout its range. Nonlinear regression analysis was used to determine fitted equations for both lines. Equations (1) and (2) show the fitted equations for the CAT and MRT, respectively. The fitted lines with the original data are shown in Fig. 3.

$$\text{CAT SCORE} = 86.72 - 4.49(\text{SNR}) - 0.44(\text{SNR})^2 - 0.0064(\text{SNR})^3, \qquad R^2 = 0.995, \quad (1)$$

$$\text{MRT SCORE} = 98.91 + 4.18(\text{SNR}) + 0.17(\text{SNR})^2 + 0.0072(\text{SNR})^3, \qquad R^2 = 0.998. \quad (2)$$

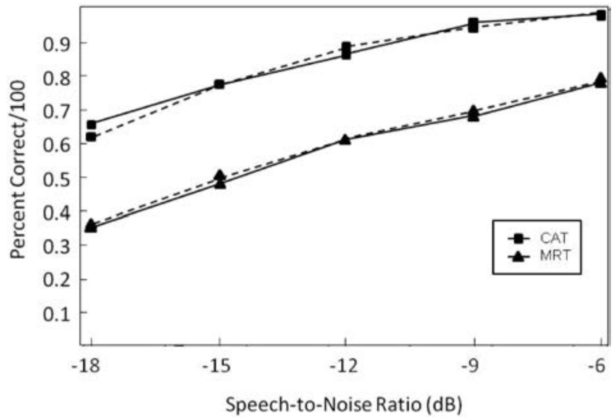

Fig. 3. Performance-intensity functions (solid lines) and regression functions (dashed lines) for CAT and MRT.

Seeing as both functions have such similar shapes and slopes, a basic model to predict CAT SI performance in 65 dB A of white noise from MRT performance data can be formulated for speech presented within a range of 47 to 59 dB A. The average difference between the SI scores at SNRs tested was 25.7%; therefore, on the basis of the data collected in this study, an upward shift of the MRT scores by 26% results in a good estimate of the CAT scores. That is,

$$\text{CAT SCORE} = 26 + \text{MRT SCORE}. \quad (3)$$

It should also be noted that the coefficients of variation for the CAT are much lower than those for the MRT (see Table 1) indicating greater repeatability of the CAT test data. However, it is important to stress that theoretical shapes of the MRT and CAT PI functions are not parallel and such approximate parallel behavior has been only assumed for practical purposes and for the limited range of SNRs investigated in this study. The MRT test is a 6-alternative test with a correct guess ratio of 1/6 (16.6%) while the CAT is a 126-alternative test with a guess ratio of 1/126 (0.8%). This difference in correct guess ratios causes the shapes of their respective PI functions to be very different at low SNR levels and the two functions can be only approximated as parallel in a relatively narrow range in mid-to-high SNRs as reported in this study. The theoretical shapes of both the MRT and CAT PI functions based on the data reported in this study are shown in Fig. 4.

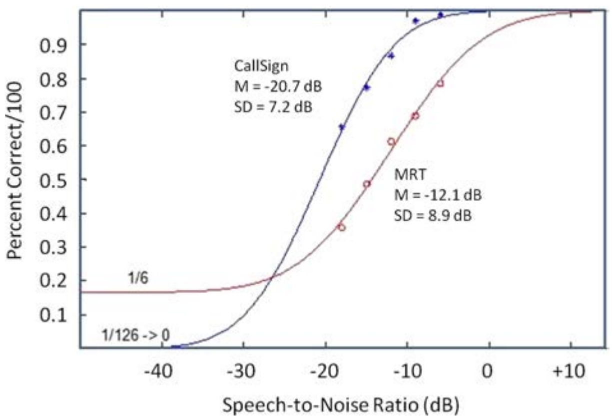

Fig. 4. Theoretical shapes of MRT and CAT PI functions derived for the data reported in the study.

The original MRT data and shifted performance-intensity functions for the CAT are shown in Fig. 5. If the MRT function was shifted to match CAT data, the shifted MRT function woud reach 100% intelligibility around −7 dB SNR, which corresponds to about 75% intelligibility for the original MRT function. Thus, the applicability of Eq. (3) under the conditions used in this study is limited to the SNRs between −18 and −6 dB. In addition, the validity of this equation may be limited to speech levels below 70 dB SPL since above this level the signal level, in addition to SNR, affects speech intelligibility (Studebaker *et al.*, 1999).

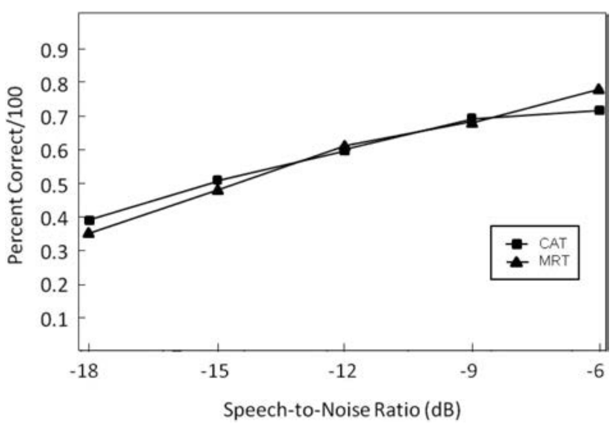

Fig. 5. Original MRT PI function and shifted CAT PI function matching MRT data.

The CAT scores obtained in this study closely agree with the data reported previously for a similar range of SNRs by Rao and Letowski (Rao, Letowski, 2003; 2006). Likewise, reported MRT scores are similar to those that would be predicted from the normal cumulative fit to the House *et al.* (1965) data as well as to those reported by Zera (2004) for a pink noise masker and Williams and Hecker (1968) for an additive speech-shaped noise. Similar data were also reported by Nickerson *et al.* (1960) for the Fairbank's Rhyme Test presented in random noise. Some small differences between data reported in our, these, and other studies result possibly from differences in masking noise and SNR measurement methods employed in these studies.

5. Summary and conclusions

The objective of the presented study was to compare the effects of SNR on the SI scores of the CAT and MRT tests conducted in 65 dB A white background noise. As expected, the results showed that both the type of speech test and SNR have significant effects on SI scores. Further analysis showed that the CAT SI scores were significantly higher and relatively less variable than the MRT SI scores for the same SNRs. In addition, the study revealed a strong positive relationship between the CAT and MRT scores across the tested range of SNRs.

Due to the fact that the MRT and CAT scores are highly correlated and both tests result in similar data variability patterns for the SNRs tested, it can be concluded that the PI functions of both tests have approximately equivalent shapes in the −18 dB to −6 dB SNR range when both tests are used in the presence of a 65 dB A white noise. By adding a 26% constant to the MRT score (Eq. (3)) we can predict the CAT score under the test conditions evaluated in the current study and maintain continuity of the data pattern using MRT at better SNRs if needed. The use of CAT in place of MRT for adverse military listening conditions below −6 dB SNR saves time and increases data repeatability. One of the important properties of the CAT is its simple vocabulary that may be easily ported to other languages. It is expected that CAT data may be relatively language independent but this concept has yet to be tested.

Acknowledgment

The authors express their gratitude to Dr. Jan Zera for providing idea for and help with Fig. 4.

References

1. Blue M., Ntuen C., Letowski T. (2004), *Speech intelligibility of the callsign acquisition test in a quiet environment*, International Journal of Occupational Safety and Ergonomics, **10**, 179–189.
2. Blue M., Ntuen C., Letowski T. (2010), *Speech intelligibility measured with shortened versions of callsign acquisition test (CAT) lists*, Applied Ergonomics, **41**, 291–294.
3. Fairbanks G. (1958), *Test of phonemic difference: The rhyme test*, Journal of the Acoustical Society of America, **30**, 596–600.
4. House A., Williams C., Hecker M., Kryter K. (1965), *Articulation testing methods: Consonantal differentiation with a closed-response set*, Journal of the Acoustical Society of America, **37**, 158–166.
5. Howes D.H. (1957), *On the relation between intelligibility and frequency of occurrence of English words*, Journal of the Acoustical Society of America, **29**, 296–303.
6. Letowski T., Karsh R., Vause N., Shilling R., Ballas J., Brungart D., McKinley R. (2001), *Human factors military lexicon: Auditory displays*, ARL Technical Report (ARL-TR-2526), U.S. Army Research Laboratory, Aberdeen Proving Grounds, MD.
7. Morton J. (1969), *Interaction of information in word recognition*, Psychological Review, **76**, 165–178.
8. Nickerson J.F., Miller A.W. Jr., Shyne N.A. (1960), *A comparison of five articulation tests*, U.S. Air Force Research and Development Command, Report No. RADC-TR-60-71, Griffiss AFB (NY): U.S. Air Force.
9. Rao M., Letowski T. (2003), *Speech intelligibility of the call sign acquisition test (CAT) for army communication systems*, Audio Engineering Society 114th Convention, Paper 5836, Amsterdam, The Netherlands.
10. Rao M., Letowski T. (2006), *Callsign acquisition test (CAT): Speech intelligibility in noise*, Ear and Hearing, **27**, 120–128.
11. Schultz M.C. (1964), *Word familiarity influences in speech discrimination*, Journal of Speech Hearing Research, **7**, 395–400.
12. Studebaker G.A. (1985), *A "rationalized" arcsine transform*, Journal of Speech and Hearing Research, **28**, 455–462.
13. Studebaker G.A., Sherbescoe R.L., McDaniel D.M., Gwaltney C.A. (1999), *Monosyllabic word recognition at higher-than-normal speech and noise levels*, Journal of the Acoustical Society of America, **105**, 2431–2444.
14. Williams C.E., Hecker H.L. (1968), *Relation between intelligibility scores on four test methods and three types of speech distortion*, Journal of the Acoustical Society of America, **44**, 1002–1006.
15. Zera J. (2004), *Speech intelligibility measured by adaptive maximum-likelihood procedure*, Speech Communication, **42**, 313–328.