Defense Technical Information Center Compilation Part Notice

ADP014026

TITLE: The Validation of Military Callsign Intelligibility

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Multi-modal Speech Recognition Workshop 2002

To order the complete compilation report, use: ADA415344

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP014015 thru ADP014027

UNCLASSIFIED

The Validation of Military Callsign Intelligibility

Celestine A. Ntuen & Misty Blue
The Institute for Human-Machine Studies
Department of Industrial & Systems Engineering
North Carolina A&T State University

Abstract

This study was conducted to evaluate the performance of human perception of speech generated by computers under normal and stressful military environments. Performance intensity (PI) functions for speech intelligibility were developed. Results are used to determine human speech awareness thresholds (SAT) for quite and noise environments.

1. INTRODUCTION

Our ability to perform tasks effectively in environments such as the battlefield, airspace management (pilots and air traffic controllers), hospitals, and manufacturing systems, depend in part oin our ability to process speech signals. Effective speech communication requires clear speaking by the talker, nonrestrictive transmission channel (medium), and good hearing and speech comprehension by the listener. These capabilities have been tested using various speech material and trained takers (speech understanding tests) or listeners (speech intelligibility tests)

One of the several methods to measure our ability to process information generated by sound or speech signals is known as speech intelligibility (Logan, Greene, & Pisoni, 1989).

Speech Intelligibility (SI) is an index for measuring the minimum absolute threshold of perceiving sound in a given environment. SI is quantitatively defined as the percentage of speech units that can be correctly identified by a listener over a given communication system in a given acoustic environment or the degree to which speech can be understood during given conditions (Letowski, Karsh, Vause, Shilling, Ballas, Brungart & McKinley, 2001). Intelligibility tests evaluate the number of words or other speech units that can be correctly identified within a controlled situation. Some examples of speech intelligibility tests are documented in ISO (1986). The relevant ones to this study are:

Diagnostic Rhyme Test (DRT): The DRT uses a set of isolated words to test for consonant intelligibility in initial position (Goldstein, 1995; Logan, Greene & Pisioni, 1989). The tests consist of 96 word pairs that differ by a single acoustic feature in the initial consonant. Word pairs are chosen to evaluate the phonetic characteristics.

Modified Rhyme Test (MRT): The MRT is an extension of DRT, tests for both initial and final consonant apprehension (Logan, Greene & Pisoni, 1989[1]). The test consists of 50 sets of 6 one-syllable words that make a total set of 300 words. The set of 6 words is played one at the time and the listener marks which word he think he hears on a multiple choice answer sheet.

Diagnostic Medial Consonant Test (DMCT): The DMCT is the same type of test as the rhyme tests described before. The material consists of 96 bi-syllable word pairs like "stopper-stocker" which were selected to differ only with their intervocalic consonant.

2. MILITARY CALLSIGN TEST (CAT)

The Auditory Research Team at the United States Army Research Laboratory developed the CAT test (Letowski, Karsh, Vause, Shilling, Ballas, Brungart, & McKinley, 2001). The CAT test utilizes military callsigns for calling phrase. A single callsign for CAT consists of a word and a number. The word is a two-syllable military alphabet code and a one-syllable number, for example, alpha 1 or bravo 2. due to their familiarity with test material and task environments. To maintain its ecological

Multi-modal Speech Recognition Workshop, June 10-12, 2002/NCAT-- Greensboro

validity, it is important to test the CAT in quiet conditions so as to establish a standard and a reference SI metric for comparison with other standard SI metrics(ISO 1986). The test material seems to be a good compromise between (1) simplicity and poor predictive value of monosyllabic signals and (2) complexity and memory load of nonsense sentences and long number sequences (Letowski, 2001).

The CAT test has been informally used by the ARL-ART in several studies but is still lacking proper validation and standardization. Such a process requires several steps that need to be completed before the final version of the test may be released. One of these steps is the standardization of SI and evaluation of the related performance intensity (PI) curve for CAT both in quiet and with background noise

3. PROCEDURE & METHODOLOGY

Participants

A group of 24 listeners between the ages of 18 and 45 participated. All listeners had pure-tone hearing thresholds better than or equal to 20dBHL at audiometric frequencies from 250Hz through 8000Hz (ANSI S3.6-1996) and no history of otologic pathology. An audiometric screening test was performed prior to participation in the study.

Each listener was seated at the listener station in a sound treated test booth using an IBM PC/586 computer and wearing TDH-39 testing earphones.All the instructions were displayed on the computer screen and the participant was able to use either the computer mouse or the computer keyboard for data input. The listener was asked to listen to the series of the CAT (military alphabet callsigns and one syllable numbers 1-8) items and identify them by pressing appropriate keys on the computer keyboard. Also, the main screen showed the display CAT test (Peak or RMS) and the signal-to-noise ratio (SNR) given by –18 dB, -12dB, -8dB, 0dB, 6dB, 12dB.

The listeners repeated the test with signal level increasing in 5dB steps until they achieve 95% or better on both tests (RMS and PEAK recordings). All the listeners' responses were stored in a file and subsequently imported into an Excel™ spreadsheet for analysis. Each listener participated in a single listening session. The session lasted about four hours and included audiometric screening, instructions, testing and several 10-15 minute long breaks.

4. SAMPLE RESULTS

The PI function showed some characteristics of logistics distributions See example in Figure 2).

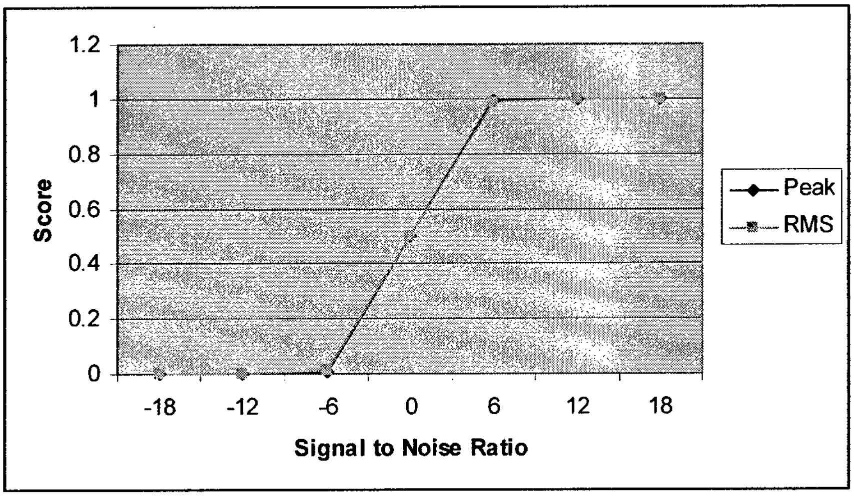

Figure 2: Sample logistics PI function for CAT intelligibility

$$Score = \frac{1}{1+e^{-0.78235*SNR}} \; ; \; R^2 = 90\%$$

(Peak) (1)

$$0 < SNR \leq 11.77$$

$$Score = \frac{1}{1+e^{-0.745*SNR}} \; ; \; R^2 = 88.24\%$$

(RMS) (2)

$$0 < SNR \leq 12.36$$

Figure 2: Sample logistics PI function for CAT intelligibility

5. CONCLUSION

The logistics PI models show that speech awareness threshold (SAT) occurs at signal-to-noise-ration (SNR) > 0, with the average listener achieving an SI value of 95% at SNR values of 11.64 for Peak and 12.22 for RMS. By using simple one parameter linear model, speech awareness threshold occurs at SNR values of approximately 2 for both Peak and RMS tests, with the average listener achieving an SI value of 95% at SNR values between 7.7 and 7.9.

References

Goldstein, M. (1995). Classification of methods used for assessment of text-to-speech systems according to the demands placed on the listener. Speech Communication, 16, 225-244.

Jekosch U. (1993). Speech quality assessment and evaluation. Proceedings of Eurospeech 93 (2): 1387-1394.

Letowski, T., Karsh R., Vause, N., Shilling, R., Ballas, J., Brungart, D., McKinley, R. (2001). Human Factors Military Lexicon: Auditory Displays. Aberdeen Proving Grounds, MD.

Letowski, T. (2001). PerformanceiIntensity function for the Callsign Acquisition Test (CAT) Research Protocol. Aberdeen Proving Grounds, MD.

Logan J., Greene B., Pisoni D. (1989). Segmental intelligibility of synthetic speech produced by rule. Journal of the Acoustical Society of America, JASA. 86 (2): 566-581.