The ability to hear clearly is critically important for the successful conduct of military operations. Unfortunately, military operations are also very noisy, and this has historically resulted in a high prevalence of hearing loss in active-duty personnel and veterans. Thus, one of the important roles of hearing surveillance programs in the military is to identify individuals with severe enough hearing impairments to endanger themselves or others when performing their mission. Most current military hearing conservation programs require annual audiometric testing to monitor Service Members for significant changes in their hearing. Those with thresholds that exceed some predefined set of frequency-dependent limits are referred to an audiologist for further testing. The results of this testing are used to make a recommendation for retention or separation. A research program is now underway to replace the current fitness-for-duty standards, which are based primarily on historical precedent, with evidence-based fitness-for-duty standards that provide a direct link between the results of clinical hearing tests and the functional hearing tasks that Service Members might need to perform on the battlefield. The proposed standards are based on the assumption that speech-in-noise perception is the task that is most likely to limit operational performance in hearing-impaired listeners. The proposed hearing threshold limits were derived from a data set containing audiometric thresholds and speech-in-noise scores for 5,487 ears. This data set was used to identify the threshold cutoffs that best identified individuals with poor speech scores without changing the total number of people referred for testing. The proposed standard also incorporates a closed-set speech-in-noise test based on the Modified Rhyme Test. We hope these new recommendations will improve the efficiency and credibility of fitness-for-duty evaluations in the military. [The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Departments of the Navy, the Army, the Department of Defense, or the US Government.]

Keywords: Fitness-for-Duty
1. Introduction

The ability to hear clearly is critically important to the safe and effective conduct of military operations, and virtually all military organizations implement some form of formal or informal regulations to ensure 1) that new recruits who are entering the military are able to hear well enough to safely execute their duties; and 2) that individuals who have acquired hearing loss during their military career can still hear well enough to complete their missions without placing themselves or others at risk from a hearing-related mishap.

In the ideal auditory fitness-for-duty monitoring program, it would be possible to identify the most essential hearing-critical tasks required from each individual Service Member and develop an objective hearing test that would directly measures the ability to perform those specific hearing-critical tasks in the exact noise environments where those tasks are executed. In the civilian and public service sectors, it is often possible to develop fitness-for-duty standards focused on a specific occupational job category, for example a police officer, corrections officer, or coast guard agent, and make a direct survey of the speech tasks and noise environments encountered in those occupations [1]. This allows a fitness-for-duty regulation to be directly tailored to the subset of hearing-critical tasks that are most relevant for those individuals. The military has a much larger range of job occupational codes, with 190 different MOS codes just for enlisted Service Members in the Army. Some of these clearly require hearing in noisy environments (MOS Field 11, Infantry) and other have a less obvious requirement for understanding speech in noisy environments (MOS Field 36, Financial Management). However, in contrast to civilian occupations, there is an expectation that all uniformed Service Members may, at some point in their careers, need to be deployed to a forward operating military base. There are also many cases where individuals in relatively low-noise occupational fields have had to cross-train to perform other operational duties that have a much greater requirement to perform hearing-critical tasks in noisy environments. This imposes a certain minimum set of hearing requirements on all Service Members, regardless of their assigned occupational specialty code. Thus, with the exception of a few specific fields with unique requirements like aviation or special operations, it is not unusual for military organizations to use the same regulations to evaluate all Service Members regardless of occupation [Note, however, that actual reassignments or separations from service are done on a case-by-case basis in a process that incorporates recommendations from the chain of command as well as the results of objective hearing tests and recommendations of the audiologist].

One common aspect of virtually all current fitness-for-duty standards is the recognition that the ability to understand speech in noise is the most essential hearing-critical task in occupations that are dangerous enough to require regular fitness evaluations. Soli et al. [1] found this to be true for law enforcement personnel and for Canadian Coast Guard officers. Bevis et al. [2] also found that 7 of the 9 hearing critical tasks identified by infantry and combat support personnel in the British military involved speech understanding in noise. Thus, one might assume that the most logical way to monitor auditory fitness-for-duty in a noise exposed population would be to conduct periodic speech-in-noise testing to ensure that nobody has experienced a shift in hearing that would make them unable to hear mission-critical voice communications. However, the current reality is that periodic hearing monitoring in the military is focused on hearing conservation, rather than auditory fitness-for-duty. This means that the infrastructure used to monitor hearing in the military is restricted to the ability to record pure-tone audiometric thresholds. Thus, the first component of any fitness-for-duty standard for the military is to identify a set of threshold values (referred to as a hearing profile) that indicates when a hearing loss has become severe enough to require an individual to receive further evaluation with a speech-in-noise test. In the next section, we describe a procedure we have used to find a set of hearing profile standards designed to use audiometric threshold data to identify individuals in the military population who are at greatest
risk for having a speech-in-noise deficit that might negatively impact their ability to perform their job duties. In Section 3, we describe a version of the Modified Rhyme Test that we have adapted to serve as a potential speech-in-noise test for the military. Finally, in Section 4, we describe the available evidence that supports the use of the Modified Rhyme Test as a general predictor of the impact that a degraded speech communication channel can have on the performance of military operations.

2. Development of Audiogram Threshold

The enrollment criteria into the hearing conservation programs varies across the different branches of the military, but all active-duty Service Members in the US Army and US Marines are required to undergo an annual hearing evaluation. The scope of these hearing conservation programs are massive, with the US Army evaluating the hearing of approximately 500,000 Service Members every year. While ideally the annual hearing evaluation would include a full audimetric work up including pure-tone thresholds, speech intelligibility in quiet and noise, and objective measures of auditory function, practical constraints on resources limits the evaluation to automated measures of pure-tone air-conduction thresholds. The pure-tone audiogram is then used to assign a hearing profile ranging from an H1 (no hearing-related operational restrictions) to H4 (hearing loss that severely limits operational capabilities and triggers medical removal from military service). The details of the threshold criteria vary across the military branches, although there are efforts to make them uniform. An H3 profile is used to identify Service Members who should undergo fitness-for-duty testing.

In recent years, there has been an effort to upgrade the current military hearing standards, which are based on medico-legal historical precedent [3], to establish a direct link between the standards and potential impairment of auditory function. The current Army threshold criteria [4] do not require unilaterally deaf individuals to undergo a fitness-for-duty evaluation, but there is a marked decrease in sound localization when the threshold at 500 Hz exceeds 40 dB HL [5]. To ensure Service Members with potential impairments to binaural function receive a fitness-for-duty evaluation, it has been proposed that Service Members with thresholds in the worse ear in excess of 40 dB HL at 500 Hz, 40 dB HL at 1000 Hz, and 60 dB HL at 2000 Hz should be given a fitness-for-duty evaluation [6].

Changes to the better-ear requirements have also been proposed to better identify Service Members with impaired speech-in-noise performance [6, 7]. Shub and colleagues [7] calculated threshold criteria based on a data set (N = 5487) that included both a measure of speech-in-noise performance and a pure-tone audiogram. The goal was to define a profiling criteria which best differentiated between individuals with normal and impaired speech-in-noise performance (defined as a score 2 standard deviations worse than the mean performance of individuals with normal-hearing thresholds) without substantially changing the total number of fitness-for-duty evaluations. The articulation index (AI) [8] or its more sophisticated successor the speech-intelligibility index (SII) [9], seem like reasonable starting points despite the standards specifically advising against using them to predict speech intelligibility for hearing impaired listeners. In addition to using the SII, Shub and colleagues [7] also examined the performance of profiling criteria based on a set of frequency-by-frequency threshold criteria. These threshold criteria were found through an iterative process that maximized the index of sensitivity given a constraint on the total number of fitness-for-duty evaluations.

Figure 1 shows the sensitivity and specificity of a few of the classifiers examined by Shub and colleagues [7]. The dashed line shows the ROC curve for a classifier based on the SII. With threshold based classifiers, there is no single criterion like there is with the SII. Rather, for a desired referral rate, the optimal frequency-by-frequency threshold criteria must be calculated. The diamonds in Fig. 1 show a set of these classifiers with the filled diamond representing the performance of the proposed rule that in which the same percentage of Service Members are referred as with the current Army rule [4] when modified to
Figure 1: The dotted line shows the ROC curve for the SII based classifier. The symbols show false alarm and hit rates for the current Army and Navy standards and the proposed optimal threshold-based classifiers.

refer individuals with suspected binaural hearing problems. This proposed rule is that a Service Member should be referred for fitness-for-duty testing if any threshold exceeds predefined limits in either the "better ear" or "worse ear", where better/worse are defined on an frequency-by-frequency basis. The criteria in the better ear are 25, 30, 25, 40, 60, and 70 dB HL at 0.5, 1, 2, 3, 4, and 6 kHz. These threshold-based classifiers outperform the SII based classifier as well as the current Army and Navy rules when modified to refer individuals with suspected binaural hearing problems.

3. Adaptation of the Modified Rhyme Test as a Speech in Noise Test

Currently, the most commonly used speech-in-noise tests for fitness-for-duty testing are the Hearing in Noise Test [10], which is used by a number of agencies to test law enforcement and public service personnel [1], and the Speech Reception in Noise Test [11], which is the current standard test for the US Army [4]. Both of these are "open set" speech tests, which require the patient to listen to speech signals presented in noise and verbally repeat them back to an audiologist, who then manually scores them as correct or incorrect. The HINT test consists of relatively simple sentences (at a 5 year old level) in speech-shaped noise, while the SPRINT consists of phonetically balanced monosyllabic words presented in a speech babble at a +9 dB SNR.

Both of the HINT and the SPRINT tests suffer from the criticism that they may not accurately represent the kinds of speech materials used in real-world occupational tasks. Although some attempts have been made to remedy this by developing speech tests with more directly-relevant speech materials, this has proven very difficult because the kinds of speech materials used in hearing-critical tasks may differ substantially even for individuals within the same military occupational specialty. An alternative approach for the development of a speech-in-noise test for fitness-for-duty evaluation is to base the test on speech materials that focus on the accurate identification of the basic phonetic elements of speech and minimize the effects that familiarity with the speech materials and/or the ability to make use of sentence context might have on the ability to perform the task. This is the approach the military has taken for the evaluation of communication systems (e.g. radios, microphones, etc) that may have to be used to deliver a variety of complex messages over a distorted radio channel and/or with talkers and listeners located in noisy environments. These tests are generally conducted with the Modified Rhyme Test (MRT) [12], a speech test where each sentence consists of a monosyllabic target word embedded in a carrier phrase of the form "You will mark <word>' please". The listener’s task is to select the target word from one of six words that differ only in terms of the leading or trailing consonant.

In linear communication systems that meet certain assumptions, mean intelligibility scores for groups
of normal-hearing subjects on phonetically-balanced CVC tests like the MRT can be predicted with reasonable accuracy by the AI [13], SII, and, for modulated noises, the more recent Extended Speech Intelligibility Index (ESII) [1]. Also, because the MRT score is monotonically related to the AI, it is possible to use MRT scores to estimate the AI and then use this estimated AI score to approximate percent correct scores with other types of speech materials that have been linked to the AI, like nonsense syllables, familiar sentences, and phonetically-balanced word lists of various lengths [13]. These relationships between the AI and scores on different speech tests is well established for group mean data on normal hearing listeners, but these group mean values may mask substantial differences in individual performance. Furthermore, although they have often been adapted to make predictions about speech in noise performance in hearing impaired listeners, the AI and SII are not considered to be valid predictors of speech intelligibility in hearing impaired listeners. However, Soli et al. [1] showed that the ESII can be used to predict the masking effectiveness of different types of modulated noise on hearing impaired listeners, thus allowing a variety of masking noises to be predicted with a speech-shaped noise masker. And one would expect the relationships between intelligibility scores on different speech materials in the same degraded listening environments to be at least partially preserved even in individual hearing impaired listeners. Thus, there is the reason to believe that results from the MRT could be generalized across other types of speech materials that might occur in military hearing-critical tasks.

3.1 Development of a Clinical Military Operational Hearing Test

The use of the MRT for speech intelligibility is outlined in ANSI/ASA S3.2-2009 (R2014), which is the American national standard method for measuring the intelligibility of speech over communication systems. In these laboratory tests, which often involve the use of a live talker and multiple live listeners, it is customary to have at least five individuals serve as talkers and for each of those five talkers to read at least three of the six 50-item balanced word lists in the MRT.

Obviously, such a comprehensive test is not suitable for use as a clinical test. Kreul et al. [14] first proposed the use of the Modified Rhyme Test as a clinical test in 1968, and they indicate they made recordings of three talkers that were calibrated to one of three speech intelligibility levels (75%, 83%, and 96%) for normal hearing listeners. Although these recordings were evaluated by a number of other researchers, the tests were never adapted for clinical use and little or no data is available on how hearing impaired listeners performed with these recordings.

In order to develop a new clinical version of the MRT, speech samples were taken from a set of MRT recordings made on six native speakers of English (3 male and 3 female) at the Air Force Research Laboratory. All of these talkers were experienced as both talkers and listeners in MRT testing, and each set of recordings included all 273 words from the MRT recorded in the "You will mark <blank> please" carrier phrase.

These recordings we processed to produce eight different listening conditions divided across three different parameters: the RMS level of the target phrase, which could be high (78 dB SPL) or low (70 dB SPL); the nominal SNR of the stimulus, which could be high (+4 dB SNR) or low (-4 dB SNR); and the binaural condition of the stimulus, which could be diotic or binaural. In the monaural conditions, the masking noise consisted of continuous Gaussian noise that was spectrally shaped to match the long-term RMS of the MRT corpus. In the binaural conditions, two independent samples of this Gaussian noise were processed to generate interaural time delays (500 μs to the left for one noise sample and 500 μs to the right for the other) and then summed together to generate the masker. In normal hearing listeners, this produced roughly a 2 dB decrease in masking relative to the diotic condition, so in the binaural conditions the actual SNR was offset 2 dB from the nominal SNR to achieve equivalent performance (i.e. the binaural stimulus were presented at -6 dB and +2 dB SNR, rather than -4 dB and +4 dB).
One goal of the clinical MRT test was to focus primarily on the distortion component of hearing loss and minimize redundancy between the MRT score and the audibility information already contained in the audiogram. In order to accomplish this, both the target speech and masking noise components of the MRT stimuli were spectrally shaped with a linear "CAM" hearing aid fit that would be appropriate for a typical hearing impaired Service Member with an H3 profile. This consisted of 20 dB gain at 500 Hz, 20 dB of gain at 1000 Hz, 35 dB of gain at 2000 Hz, and 70 dB of gain at 4000 Hz. The stimuli were also sharply low-pass-filtered at 8 kHz. This shaping occurred prior to the adjustment of the level of the target speech.

The processed MRT recordings were originally used to develop two 104-word lists, with a full factorial design of 52 trials each at high and low stimulus levels, high and low SNRs, and diotic and binaural masking conditions. These 104-word lists were later downselected into 80-word lists in order to shorten the test and optimally balance difficulty across the 8 conditions.

### 3.2 Collection of Normative Data with the MRT test

The 104-word (and later 80-word) MRT test was programmed into a tablet-based interface and used to collect normative data on Service Members who were receiving annual hearing tests at a number of different DoD hearing conservation clinics. Figure 2 shows cumulative distributions of the percentage of correct responses on the 80-word MRT tests for 2037 active duty Service Members. The Service Members have been divided into those with normal hearing (<20 HL at all frequencies from 500 Hz to 6 kHz), those with a mild hearing loss but still and H1 hearing profile according to Army Regulation 40-501, those with a moderate hearing loss (H2 profile), and those with a more severe hearing loss (H3). The horizontal line at 0.5 intersects with the median MRT score for each hearing category, which ranged from 72% for the H3s to 81% for the normal hearing listeners. The vertical dashed line is the 5th percentile line for normal hearing listeners. Roughly 40% of the listeners with H3 profiles fell in the bottom 5th percentile of normal on the MRT. This is roughly similar to the number of similarly hearing impaired listeners who fall in the bottom 5th percentile of other tests, such as the QuickSIN and the SPRINT.

### 4. Establishing a minimum performance requirement for the MRT

One of the most challenging components of establishing any auditory fitness-for-duty standard is determining a minimum level of performance on a clinical speech-in-noise test that would correspond to
what line commanders would consider to be a minimum level of acceptable performance (or, conversely, a maximum acceptable risk of inadequate performance) in a real-world mission critical task.

Because the MRT is so widely used in the military, both as a research tool and as a way for evaluating speech intelligibility with communication systems, we believe there is more data supporting a minimum standard of performance for the MRT than for any other speech test.

One of the most direct sources for establishing a minimum standard based on the MRT is in MIL-STD-1474.E, which defines the DoD Design Criteria for Noise Limits. This standard mandates that the "Sound pressure levels at the listener's positions shall not degrade speech communication performance to less than 80 percent correct, adjusted for guessing, as measured using the modified rhyme test (MRT) in accordance with ANSI/ASA S3.2 with talkers and listeners in representative ambient noise environments as specified by the procuring activity. The specified representative ambient noise environments should represent the worst nominal case where communications are required. This requirement shall apply to all communication environments, including face-to-face, public address-loudspeaker, and/or with communication headsets as specified by the procuring activity".

This standard clearly establishes 83% (80% adjusted for guessing) as the de-facto minimum acceptable level of MRT performance for conducting military operations. The standard also establishes the MRT, conducted on listeners with normal hearing, as the preferred method for determining if a given noise environment meets this 80% standard.

Although it wasn’t necessarily designed with this goal in mind, it is an exceptionally convenient accident that the clinical MRT test described in the previous section has a mean percent correct value of 80% for normal hearing listeners. This means that the composite set of listening conditions covered by the eight conditions of the MRT can be viewed as equivalent to a listening environment where normal hearing listeners could be expected to perform just at the minimum acceptable level in a speech perception task. Any drop in performance below this level, beyond what would be expected in intersubject variation in the normal population, means that one would expect the hearing impaired listener to be performing at a militarily unacceptable level in a noise situation where normal hearing listeners would be performing just at the bottom range of what is considered militarily acceptable. Thus, one could easily argue that the 5th percentile normal hearing listener score on an MRT test where the average normal hearing listening score is 80% represents a reasonable cutoff for acceptable performance on that test. In the current version of the MRT, that would be approximately 70%.

There are also a number of other studies that have shown that MRT scores less than 70% tend to be correlated with unacceptable performance in MRT tasks. Peters and Garinther [15] found roughly a 10% increase in total kill time in an M1A1 tank simulator when the MRT score dropped from 73% to 52%. Keller et al. [16] and Sheffield et al. [17] showed that performance in a simulated Naval CIC begins to decline rapidly when speech intelligibility was limited to 70% or below on the MRT.

In a separate preliminary study looking at a collaborative task between paired participants using a collaborative and interactive voice communication task called the DiaPix, it took a minute longer for on average (N=10 pairs) to find 8 differences when the individual SNR for each participant was systematically controlled to fix speech intelligibility at 70% as measured by an adaptive MRT as compared to a quiet condition.

Taken together, these studies appear to corroborate the requirement in MIL-STD-1474 for a minimum speech intelligibility level of 80% on the MRT in order to achieve adequate performance in military environments. This seems to support the use of a clinical test based on the MRT as a metric for evaluating fitness-for-duty in military personnel. Work is now underway to tie together the audiometric threshold discussed in Section 2 with a performance criterion based on the clinical MRT test in Section 3 to establish a procedure for making recommendations about auditory fitness-for-duty in hearing impaired Service Members.
REFERENCES


