NTIA Report 13-493

Intelligibility of the Adaptive Multi-Rate Speech Coder in Emergency-Response Environments

David J. Atkinson
Stephen D. Voran
Andrew A. Catellier

report series

U.S. DEPARTMENT OF COMMERCE • National Telecommunications and Information Administration

NTIA Report 13-493

Intelligibility of the Adaptive Multi-Rate Speech Coder in Emergency-Response Environments

David J. Atkinson
Stephen D. Voran
Andrew A. Catellier

U.S. DEPARTMENT OF COMMERCE

December 2012
Revised March 2013

PREFACE

It is anticipated that a new broadband public safety network will employ long term evolution (LTE) radio technology and that mission-critical voice communications will thus employ voice over LTE (VoLTE) using, as one option, the Adaptive Multi-Rate (AMR) speech coder. Our previous investigation provided quantitative confirmations of field reports that some of the unique environmental conditions associated with firefighting are problematic for digital speech coding algorithms. That investigation is presented in a 2008 NTIA Technical Report titled "Intelligibility of Selected Radio Systems in the Presence of Fireground Noise: Test Plan and Results," available at www.its.bldrdoc.gov. Those results motivate us to test the AMR speech coder for use in mission-critical public safety voice communications. Specifically, we are concerned with speech intelligibility, and the tool we have selected to assess intelligibility is the Modified Rhyme Test (MRT). Much of the work described here builds on and parallels our earlier work described in the 2008 report mentioned above.

Funding for the work described in this report was provided by the Department of Homeland Security Office for Interoperability and Compatibility and the National Institute of Standards and Technology Office of Law Enforcement Standards. The work was conducted at the Institute for Telecommunication Sciences under the supervision of Jeffrey Bratcher.

The authors wish to thank to Kathy Mayeda, Bill Ingram, and Joel Dumke of NTIA/ITS. Kathy arranged for the travel of public safety personnel that were MRT subjects. Bill and Joel administered the MRT to those subjects.

The authors also extend thanks to the following organizations, whose help was instrumental in the conduct of this experiment:

- The City of Boise, Idaho, and the Town of Plainfield, Indiana, for providing opportunities to create high-quality recordings of public safety noise environments used in this work and/or other related work.

- The following agencies and jurisdictions for adjusting schedules and allowing their public safety personnel to travel to Boulder to participate in the MRT described in this report:
  - Beavercreek Fire Department, Dayton, OH
  - Boise Fire Department, Boise, ID
  - DC Fire & EMS Department, Washington, DC
  - Engine Company 24, #2 Platoon, Washington, DC
  - FBI National Academy Association, Orlando, FL
  - Houston Fire Department, Houston, TX
  - Meridian Fire Department, Meridian, ID
  - Millstone Valley Fire Department, East Millstone, NJ
  - New Jersey Police Task Force, Newark, NJ
  - Seattle Police Department, Seattle, WA
  - Somerset Fire Company, Somerset, NJ
  - State of Michigan, Grand Rapids, MI
  - US Army Criminal Investigation Labs, Atlanta, GA

    - Washington DC Fire Department, Baltimore, MD

- The following companies who loaned equipment or provided support for equipment used in this work and/or other related work.
    - Scott Health and Safety
    - Motorola
    - Draeger Safety
    - EF Johnson
    - International Safety Instruments

Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS AND ACRONYMS

ACELP	algebraic code-excited linear prediction
AMBE™	Advanced Multi-Band Excitation (a trademark of Digital Voice Systems, Inc.)
AMR	Adaptive Multi-Rate
AMR7.4	AMR speech coding using the 7.5 kb/s mode
AMR12.2	AMR speech coding using the 12.2 kb/s mode
Analog	25 kHz bandwidth FM land mobile radio link
b	bits
CLI	command line interface
dB	decibel
dB HL	decibel, hearing level
dBA	decibel, A-weighted
dBC	decibel, C-weighted
DSP	digital signal processing
FEC	forward error correction
HATS	head and torso simulator
LRP	lip reference point
LTE	long-term evolution
MRT	Modified Rhyme Test
ms	millisecond
OBSA	Octave Band Set A
OBSB	Octave Band Set B
P25	full rate AMBE+2™ speech encoder and decoder, version 1.6, as would be found in P25 digital radio pairs
PASS	personal alert safety system
s	second
SCBA	self-contained breathing apparatus
smpl	samples
SNR	signal-to-noise ratio
VoLTE	voice over long-term evolution

EXECUTIVE SUMMARY

Recent U.S. legislation has paved the way for the creation of the "First Responder Network Authority," or FirstNet, to oversee the creation of a new nationwide interoperable broadband network supporting the work of public safety officials. It is anticipated that this broadband public safety network will employ long term evolution (LTE) radio technology, and that traditional telephony voice communications could be accomplished through voice over LTE (VoLTE). A popular choice for digital speech coding in VoLTE systems is the Adaptive Multi-Rate (AMR) speech coder. This speech coder delivers very good speech quality and very high speech intelligibility in traditional wireless telephony applications.

A previous investigation confirmed field reports that some of the unique environmental conditions associated with firefighting can be problematic for digital speech coding algorithms [1]. When high noise levels are present (e.g., from pumps, nozzles, saws, or alarms) the signal presented to a digital speech coder can deviate significantly from what the speech coder anticipates, and as a result output speech can become distorted. In extreme cases, the intelligibility of the output speech may suffer. This problem can be compounded when the mask associated with a self-contained breathing apparatus (SCBA) is used because a mask can further alter the speech signal. This may cause additional difficulties for the speech coder and associated distortion and reduction of intelligibility.

If the AMR speech coder is to be used in emergency-response environments, it is important to investigate its robustness to the problems described above. The work presented here addresses this question: How does the speech intelligibility of the AMR speech coder compare with the speech intelligibility of analog and digital reference systems when used in emergency-response environments? For this work the analog reference system is the time-honored 25 kHz FM radio link. The digital reference system is a Project 25 digital radio link. We evaluated AMR coders operating in the 12.2 kbit/s mode and the 7.4 kbit/s mode.

In collaboration with our public safety partners, we have selected seven important environments for this investigation: five of these include masks, and four include background noises. The testing protocol was also selected in collaboration with these partners. It starts with a standardized protocol used for testing "face-to-face" voice communication through an SCBA mask. We extended the specified mouth, mask, ear communication path to form a mouth, mask, communication system, ear communications path. We used sound-isolated booths, head and torso simulators, and various types of studio quality audio equipment to produce recordings simulating the relevant key aspects of the environments and systems under test.

We evaluated the intelligibility of the resulting recordings by means of the Modified Rhyme Test (MRT). In an MRT trial, a subject must identify the word he heard from a set of six words that rhyme. For example, the subject is presented with a recording that says "Please select the word bed." Six options are presented through a graphical user interface (GUI) on the screen of a handheld device. For this example those options are: "led," "fed," "bed," "red," "wed," and "shed." If the recording has high intelligibility, the word "bed" will be clear and will be correctly selected in a majority of the trials. If the recording has low intelligibility, subjects may select other, incorrect, options based on what they thought they heard. Fifteen subjects participated in the MRT and all were public safety practitioners.

Our results are based on the analysis of 26,900 MRT trials. The relative speech intelligibility of the four systems evaluated (denoted AMR12.2, AMR7.4, Analog, and P25 for brevity) is strongly influenced by the choice of operating environment. Depending on which of the environments is considered, AMR12.2 speech intelligibility may be above, the same as, or below Analog intelligibility, but the AMR12.2 intelligibility is always above or the same as P25 intelligibility, regardless of environment. AMR7.4 speech intelligibility may be the same as or below the Analog intelligibility depending on the environment. AMR7.4 intelligibility is above or the same as P25 intelligibility in all environments except the environment with no noise and no mask.

INTELLIGIBILITY OF THE ADAPTIVE MULTI-RATE SPEECH CODER IN EMERGENCY RESPONSE ENVIRONMENTS

David J. Atkinson, Stephen D. Voran, and Andrew A. Catellier[1]

This report describes speech intelligibility testing conducted on the Adaptive Multi-Rate (AMR) speech coder in several different environments simulating emergency response conditions and especially fireground conditions. The intelligibility testing protocol was the Modified Rhyme Test (MRT). Conditions included background noises of various types, as well as a mask associated with a self-contained breathing apparatus. Analog FM radio transmission and Project 25 digital radio transmission were also included in the test as reference points. Test participants were persons employed as first responders in public safety fields. Through statistical analysis of 26,900 MRT trials we are able to draw conclusions on speech intelligibility for AMR speech coding relative to analog and digital radio reference points for five different operating environments.

Key words: AMBE; AMR; analog FM; LMR; fireground; MBE; MRT; P25; public safety; SCBA mask; speech coder; speech intelligibility; subjective testing;

1 INTRODUCTION

Public safety first responders have relied on two-way radios to communicate mission-critical voice messages for many decades. Over the years these radios have evolved to provide greater reliability and flexibility. Analog FM radios were the standard through the 1980s.

In 1989, Project 25 was initiated in order to develop a standardized digital radio and to promote interoperability among digital land mobile radio (LMR) systems. Project 25 is a cooperative agreement between the Association of Public Safety Communications Officials (APCO) and the Telecommunications Industry Association. The result is a highly successful, widely deployed digital radio standard serving public safety first responders across the U.S. Digital speech coding in Project 25 radios is accomplished through Advanced Multi-Band Excitation (AMBE™, a trademark of Digital Voice Systems, Inc.) as described in [2] and [3]. Due to the narrowband nature of the system, speech is encoded at relatively low bit rates (e.g., 2450 and 4400 bps).

In 2012, U.S. legislation paved the way for the creation of the “First Responder Network Authority,” or FirstNet [4]. This independent body will oversee the creation of a new nationwide interoperable broadband network supporting the work of public safety officials on behalf of the U.S. Department of Commerce’s National Telecommunications and Information Administration. It is anticipated that this broadband public safety network will employ long term evolution (LTE) radio technology, and that traditional telephony voice communications could be accomplished

[1] The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.

through voice over LTE (VoLTE). Because LTE is a broadband network, speech coding at higher bit rates is possible. A popular choice for VoLTE is the Adaptive Multi-Rate (AMR) speech coder. AMR operates at a range of bit rates from 4.75 to 12.2 kbps and is already well-established in 3G and 4G wireless services [5],[6].

A previous investigation confirmed field reports that some of the unique environmental conditions associated with firefighting are problematic for digital speech coding algorithms [1]. This is an understandable limitation. Efficient, low bit rate digital speech coding is possible because speech signals are a specific category of audio signals with well-known statistical properties. When background noise is added to a speech signal, or the speech is distorted in other ways, the resulting input signal statistics can deviate significantly from the anticipated statistical properties and thus the output signal can become distorted. In extreme cases, the intelligibility of the output speech may suffer. Robustness to background noise and distorted speech is a goal for every speech coder. However, this goal must compete with other important goals including high quality for pure speech, low bit rate, low complexity, low delay, and robustness to channel losses.

This motivates us to test the AMR speech coder for use in emergency-response environments. More specifically, we seek to answer this question: *How does the speech intelligibility of the AMR speech coder compare with the speech intelligibility of analog and digital reference systems when used in emergency-response environments?* The analog reference system is a 25 kHz FM radio link and the digital reference system is a Project 25 digital radio link.

Emergency-response environments are indeed unique, and also diverse. In collaboration with our public safety partners, we have selected seven important environments for this experiment, with a particular emphasis on fireground environments. Our testing protocol was also selected in a collaborative fashion. It starts with the protocol described in [7] for testing "face-to-face" voice communication through the mask associated with a self-contained breathing apparatus (SCBA) in noisy environments. To quantify communications systems, we extended the mouth, mask, ear communication path laid out in [7] into a mouth, mask, communication system, ear communication path.

After [7] specifies much of the testing environment, it then specifies the Modified Rhyme Test (MRT) as the actual testing mechanism. The MRT is fully defined in [8]. In an MRT trial, a subject must identify the word he heard from a set of six words that rhyme. As in our earlier work, we use public safety first responders as test subjects [1].

Section 2 provides descriptions of four different communications systems under test and seven operating environments. Two systems were AMR based, and two were the analog and digital reference systems. Operating environments were defined by background sounds that may have been present, a mask associated with an SCBA that may have been present, and the position of a microphone relative to any present mask. In Section 3 we detail the production of recordings used in the MRT. Standardized lists of rhyming words were recorded and then processed to simulate the 28 combinations of communications systems and environments. Further details of the MRT are covered in Section 4, including the physical environment for the test, the details of a single trial, the overall test structure, and information about the test subjects and their participation. Section 5 provides the results obtained through statistical analysis of 26,900 MRT trials. Finally, we draw conclusions regarding the relative performance of the four systems.

2 COMMUNICATION SYSTEMS AND ACOUSTIC ENVIRONMENTS

This section describes the four communications systems (or portions thereof) under test. It also describes the seven acoustic environments used in this test. The test combined all four communications systems with all seven acoustic noise environments to create a total of 28 test conditions. Of these 28 conditions, 24 produced usable MRT results. A defective amplifier was in the signal path for four of the 28 conditions and this was discovered only after testing was completed. More details are provided in Section 3.2.

2.1 Communication Systems

Four communication systems were considered in this test:

- **System 1.** Software simulation of analog FM radio pair using 25 kHz wide channel with. receive radio in the full quieting state. Denoted by "Analog" for brevity at numerous places in this report.

- **System 2.** Full rate AMBE+2™ speech encoder and decoder software with enhancements (version 1.6), as would be found in P25 digital radio pairs currently (2012) on the market. The speech coder uses 4400 b/s for speech coding and 2800 b/s for forward error correction (FEC), resulting in a gross bit rate of 7200 b/s. No added radio channel impairments. Denoted by "P25" for brevity at numerous places in this report.

- **System 3.** AMR speech encoder and decoder software using 12.2 kb/s for speech coding as might be used in VoLTE under most radio conditions. No added radio channel impairments. Denoted by "AMR12.2" for brevity at numerous places in this report.

- **System 4.** AMR speech encoder and decoder software using 7.4 kb/s for speech coding as might be found in VoLTE under difficult radio conditions. No added radio channel impairments. Denoted by "AMR7.4" for brevity at numerous places in this report.

System 1 is the time-honored reference system that public safety practitioners have used for decades. System 2 is the newer, digital reference system, first deployed in the 1990s and now widely used. Systems 3 and 4 are those that we wish to evaluate, especially relative to the analog and digital reference systems.

The AMR speech coder offers speech coding at eight different bit rates (i.e., 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2, and 12.2 kbps) and can change its rate at the start of any 20 ms speech frame [5],[6]. When radio conditions deteriorate, the speech coding rate can be reduced, thus allowing for the addition of more robust channel coding. The goal is to allow a call to continue, but with lowered speech quality. This is a desirable alternative to unintelligible speech, or a dropped call.

The underlying core speech coding algorithm used in AMR is called algebraic code-excited linear prediction (ACELP). Linear prediction provides a popular and efficient speech representation consisting of a linear filter and an excitation signal. In ACELP speech coding, the excitation signal is efficiently constructed from a set of sparse, strategically organized codewords stored in a set of codebooks. AMR speech coding is already well-established in 3G and 4G wireless services.

2.2 Transmit Side Acoustic Environments

Seven acoustic environments were considered for the transmit location. Acoustic environments were defined by any noise that was added, any SCBA mask present, and how any such mask was used with the communication system. The added noises covered the range from best-case (no noise added) to a very difficult environment (chainsaw noise added at 5 dB SNR):

- **Environment 1.** No noise added. The only noise present in the transmitted speech was the unavoidable very low level noise associated with recording processes.
- **Environment 2.** Recorded nightclub noise added at 5 dB SNR. This noise included a rock band and a talkative crowd.
- **Environment 3.** No noise added. SCBA mask present, speech captured by microphone at vox port (also called mechanical voice diaphragm).
- **Environment 4.** No noise added. SCBA mask present, speech captured by microphone internal to mask.
- **Environment 5.** Personal alert safety system (PASS) alarm sounding, SNR -2 dB. SCBA mask present, speech captured by microphone at vox port.
- **Environment 6.** PASS alarm sounding, SNR -2 dB. SCBA mask present, speech captured by microphone internal to mask.
- **Environment 7.** Recorded chainsaw noise added at 5 dB SNR. SCBA mask present, speech captured at vox port. In the noise recording, the chainsaw was cutting a wooden roof.

Power spectral densities and spectrograms for some of the noise environments are available in [1]. The SNRs listed above describe the environment before any mask effects. SCBA masks can change the relative levels of the wearer's speech and any background noise that may be present, and thus they can change the effective SNR at the input to the transmitting device. Masks also change the timbre or tonal balance of background noise and the wearer's speech. Both the SNR and timbre depend on the microphone placement (internal to mask or at mask vox port) as well. Both the SNR and the timbre can impact the intelligibility of the speech at the receiver.

In Section 3.2 we describe how the seven different environments were created for the different systems used in this test.

2.3 Receive-Side Acoustic Environment

Our testing approach was motivated by the protocol described in [7] for testing "face-to-face" voice communication through SCBA masks in noisy environments. That protocol specifies a noise environment in which the listening subject must perform the MRT tasks: pink noise covering the band from 400 to 4000 Hz. Thus, this became the specification for the receive-side acoustic environment in our test, i.e., the acoustic environment of the listening lab where the MRT is conducted. In Section 4.1 we describe the configuration of this listening lab, and the processes by which we created and verified this pink noise environment.

3 PREPARATION OF RECORDINGS FOR MODIFIED RHYME TEST

3.1 Production of Original Speech Recordings

The MRT protocol specifies sets of rhyming words [8]. There are 50 sets and each set contains 6 words. Some of the sets contain words that rhyme in the strict sense, for example “bed,” “led,” “fed,” “red,” “wed,” and “shed.” Other sets contain words that rhyme in the more general sense—the words display some type of phonetic similarity. An example is the set “dug,” “dung,” “duck,” “dud,” “dub,” and “dun.” In the MRT, each word was presented in a carrier sentence: “Please select the word —.” For example, when the test word was “bed,” the carrier sentence was “Please select the word bed.”

Two female and two male talkers were used to record the MRT words in the carrier sentence. Each was a native speaker of North American English with no distracting accent. Each talker recorded 300 sentences, consisting of the 50 sets of 6 words, each in the standard carrier sentence. This is a total of 1200 recorded sentences.

The recordings were made using high-quality audio equipment in a quiet environment. The recording room was a sound-isolation chamber with a noise criterion rating of NC-35 [9]. We used a Shure Beta 53A microphone sampled at 48,000 smpl/s, 16 b/smpl for direct-to-disk recording in monaural .wav format files. Care was taken to avoid clipping and low signal levels. After recording was completed, the active speech level for each sentence was normalized to 28 dB below overload using the ITU-T Recommendation P.56 voltmeter software [10],[11].

3.2 Processing of Original Speech Recordings to Produce MRT Recordings

The next step was to process the original speech recordings in order to produce the MRT recordings. This involved simulating the elements of the environment (noises, mask, and microphone location) and simulating the relevant portions of the communication system. With four systems and seven environments to test, twenty-eight different recording configurations were required. Each of the 1200 recorded sentences was passed through each of these 28 recording conditions to produce a total of 33,600 MRT recordings.

The recording configurations shared some common elements. One such element was a sound-isolated booth. This booth provided a controlled acoustic environment. Loudspeakers inside the booth could reproduce a desired background noise field and that field was not disrupted by noises produced outside the booth. Sound isolation works in both directions, so experimenters outside the booth were protected from exposure to noise fields that may have been distracting or dangerous. An example photo of a sound-isolated booth is shown in Figure 1.

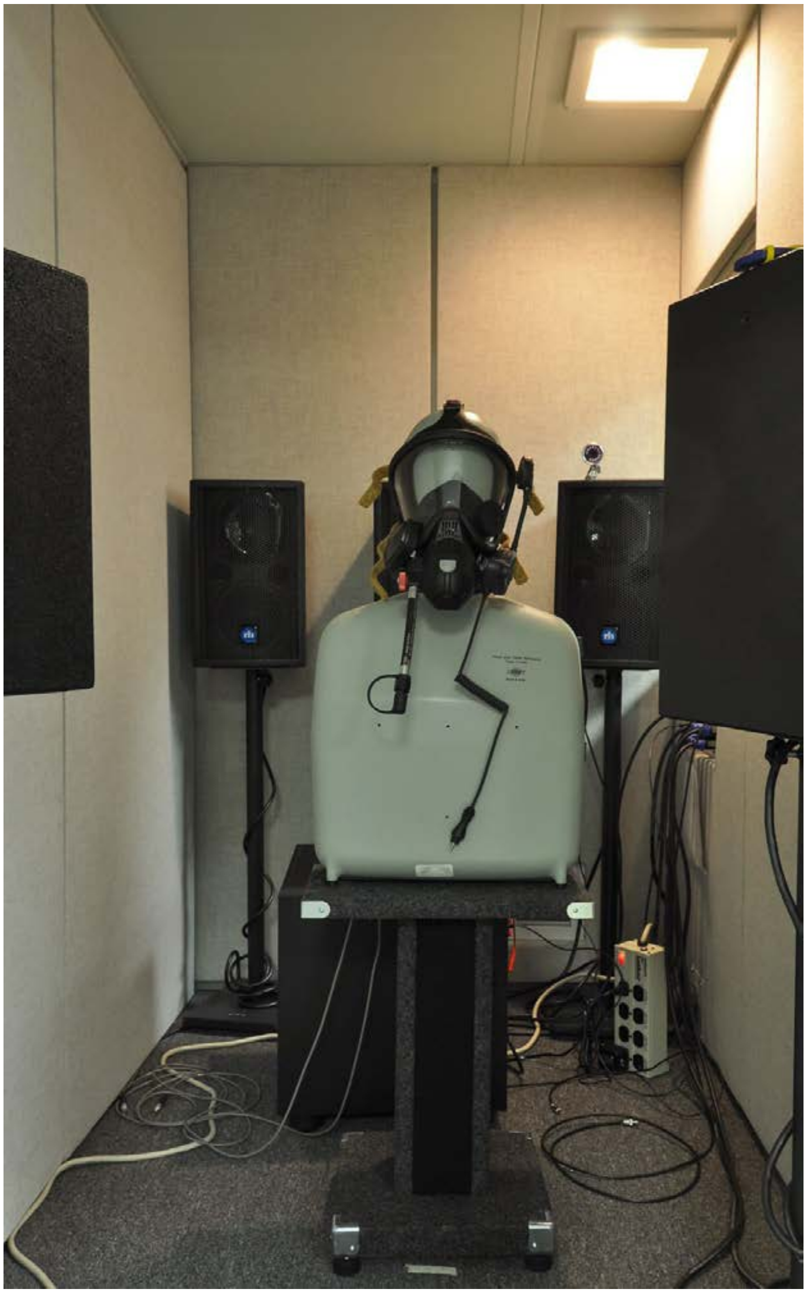


Figure 1. HATS set up in a sound-isolated booth.

A head and torso simulator (HATS) [12] is a precision crafted waist-up manikin that can simulate relevant acoustic properties of a human engaged in typical uses of a communication system. In talking mode, a small, calibrated, precision loudspeaker located inside the head of the manikin caused sound to radiate from the lips of the HATS, accurately simulating the way sound radiates from the lips of a human when talking. Figure 2 shows a HATS in this configuration. Figure 3 shows a HATS in talking mode and wearing a mask with a microphone set up to receive an audio signal from the mask’s vox port. Icons that represent three HATS configurations are shown in Figure 4.

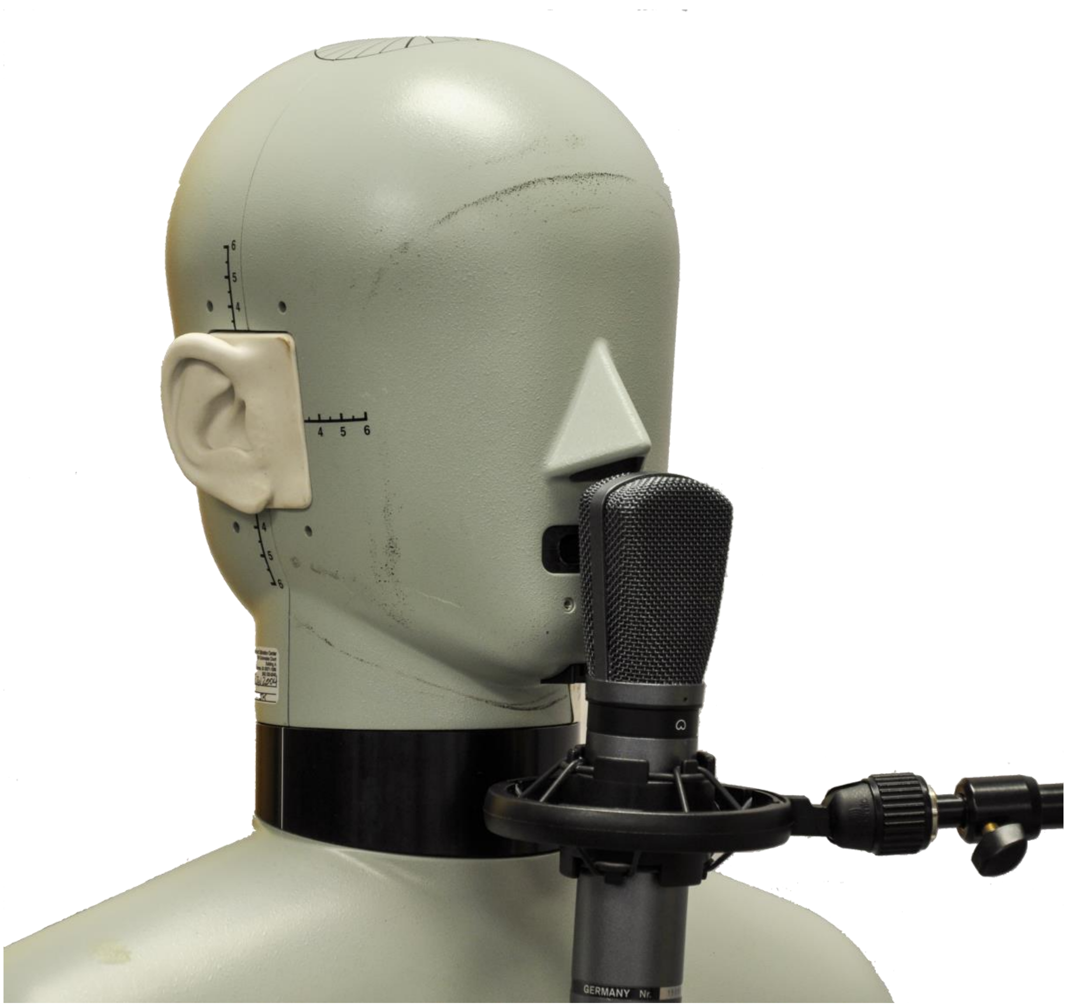


Figure 2. Photograph depicting a HATS speaking directly to a microphone.

Figure 3. Photograph depicting a HATS set up to speak through the vox port of an example SCBA mask to a microphone.

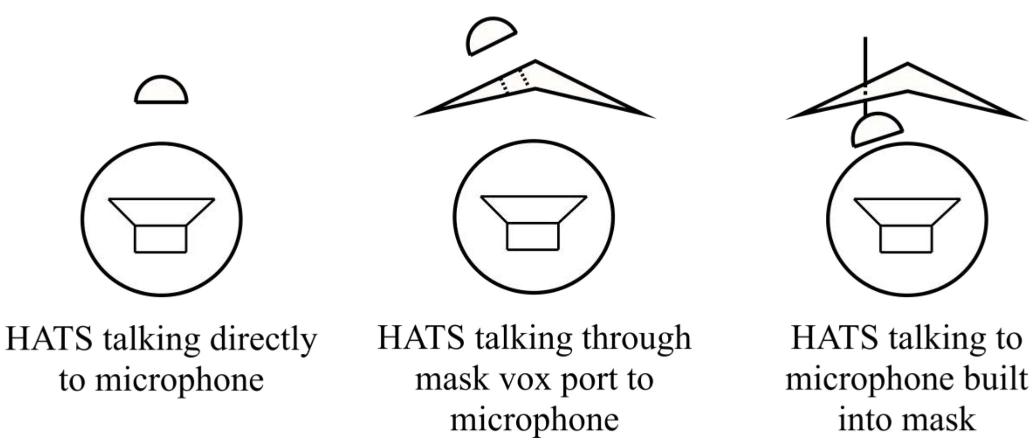


Figure 4. Icons used to represent HATS configurations in subsequent diagrams.

Studio-quality audio equipment was required to play, record, and control the various signals involved in the recording configuration. One playback path is the path from a digital audio file to the analog input of the HATS loudspeaker. Another playback path runs from a digital audio file to the analog inputs of the loudspeakers that produce the background noise field in the booth. One recording path was the path from a microphone to a digital audio file.

Figure 5 shows the configuration used to adjust speech and noise levels. A calibrated microphone was placed 2.5 cm (1 inch) from the lip reference point (LRP) of the HATS. The level of the signal driving the loudspeaker inside the HATS was adjusted to obtain a level of 100 dBC at the microphone location with no background noise present. If the environment was one that included noise, then the speech signal was muted and the level of the signal driving the noise loudspeakers was adjusted to obtain the SNR specified for the environment. For example, in Environment 2 a 5 dB SNR is specified, so the noise level was adjusted to obtain a reading of 95 dBC.

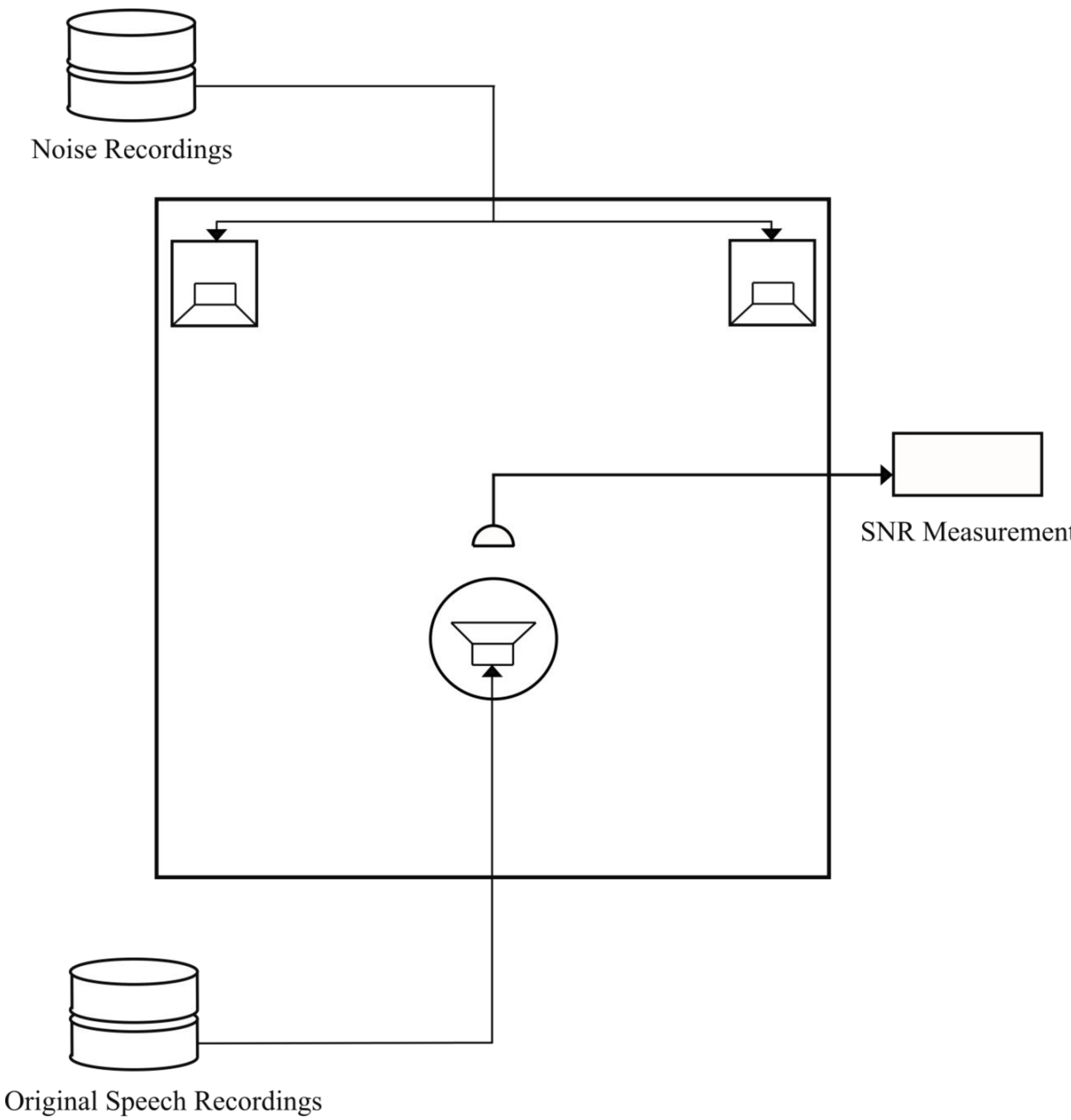


Figure 5. Block diagram showing how SNR measurements were made.

All four systems were implemented in software. Effects of impaired radio channels are outside the scope of the present investigation. Thus for the analog system the software simulated the receiving radio in the full-quieting state. The remaining systems used digital speech coding. No bit-errors were introduced and no data was deleted. Each speech decoder received the exact bit stream produced by the corresponding speech encoder.

Figure 6 shows the recording configuration for Environments 1 and 2. The microphone was positioned 2.5 cm (1 inch) from the talking HATS LRP. The original speech recording then propagated from the lips of the talking HATS into the microphone and was digitized and stored, then passed through the appropriate software to simulate the desired system, then stored again as an MRT recording. In Environment 1 this was done without the addition of any noise. In Environment 2, recorded nightclub noise (with level adjusted for 5 dB SNR) was played through the noise loudspeakers.

Figure 7 shows the recording configuration for Environments 3, 5, and 7. Here the talking HATS wore an SCBA mask and the microphone was located immediately next to the vox port (also called the mechanical voice diaphragm) on that mask. No noise was added for Environment 3. A PASS alarm recording was added in Environment 5, and the chainsaw recording was used in Environment 7.

Figure 8 shows the recording configuration for Environments 4 and 6. This is similar to the previous figure, but shows that the acoustic signals were captured by a microphone located inside and integral to the mask. The microphone signal then passed through a battery-powered amplifier that was also integral to the mask. No noise was added for Environment 4, and the PASS alarm recording was used in Environment 6.

Recordings for Environments 4 and 6 were acquired at three different times. Initial recordings were made for the Analog and P25 systems. It was then determined that the integral amplifier in the mask was defective in a manner that caused attenuation of high-frequency speech content. The defective amplifier was replaced. In a second step replacement recordings were made for the Analog and P25 systems. In a third step recordings for AMR12.2 and AMR7.4 were made. In the process of organizing the 33,600 speech files for the MRT an error was introduced. The intent was to use the recordings from the second and third steps, but in fact the recordings from the first and third steps were used. The recordings from the first step included the defective amplifier in the speech path. The result is that for Environments 4 and 6, the recordings for the Analog and P25 systems were not valid. Thus, this report does not analyze the results for these four conditions. The total number of valid conditions is 24 rather than 28.

System 1 is a software simulation of 25 kHz analog FM. The software includes representative filters as would be used in typical analog transmitters and receivers, as well as frequency modulation, deviation limiting, and frequency demodulation. The software meets the specifications for audio filtering and deviation limiting given in the TIA-603 standard for 25 kHz analog channels. While this is representative of typical radio designs, alternative designs are also possible.

The software for the P25 reference (System 2) was provided under license by Digital Voice Systems, Inc. It implements the full rate AMBE+2™ speech encoder and decoder with enhancements. This Windows® command line interface (CLI) executable development software is labeled "version 1.40e, October 14, 2009" but is equivalent to the DSP production code version 1.6, as would be found in P25 digital radio pairs currently (2012) on the market. "Full rate" indicates 4400 b/s for speech coding and 2800 b/s for FEC, resulting in a gross bit rate of 7200 b/s.

For AMR speech coding (Systems 3 and 4) we used the floating point Windows CLI executable AMR implementation provided by VoiceAge Corporation through the Open AMR Initiative. This software is compliant with [5] and is available at http://www.voiceage.com/freecodecs.php.

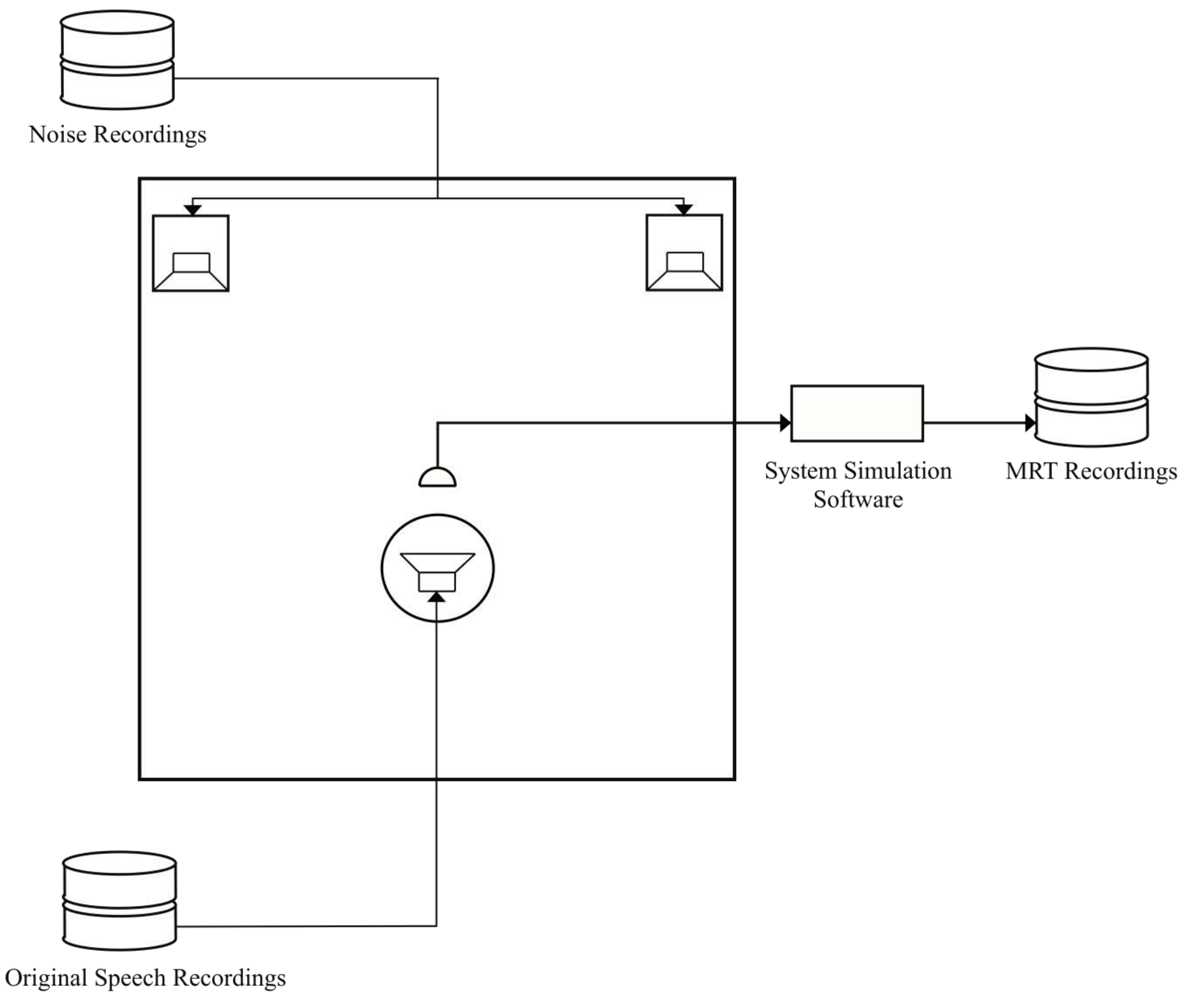
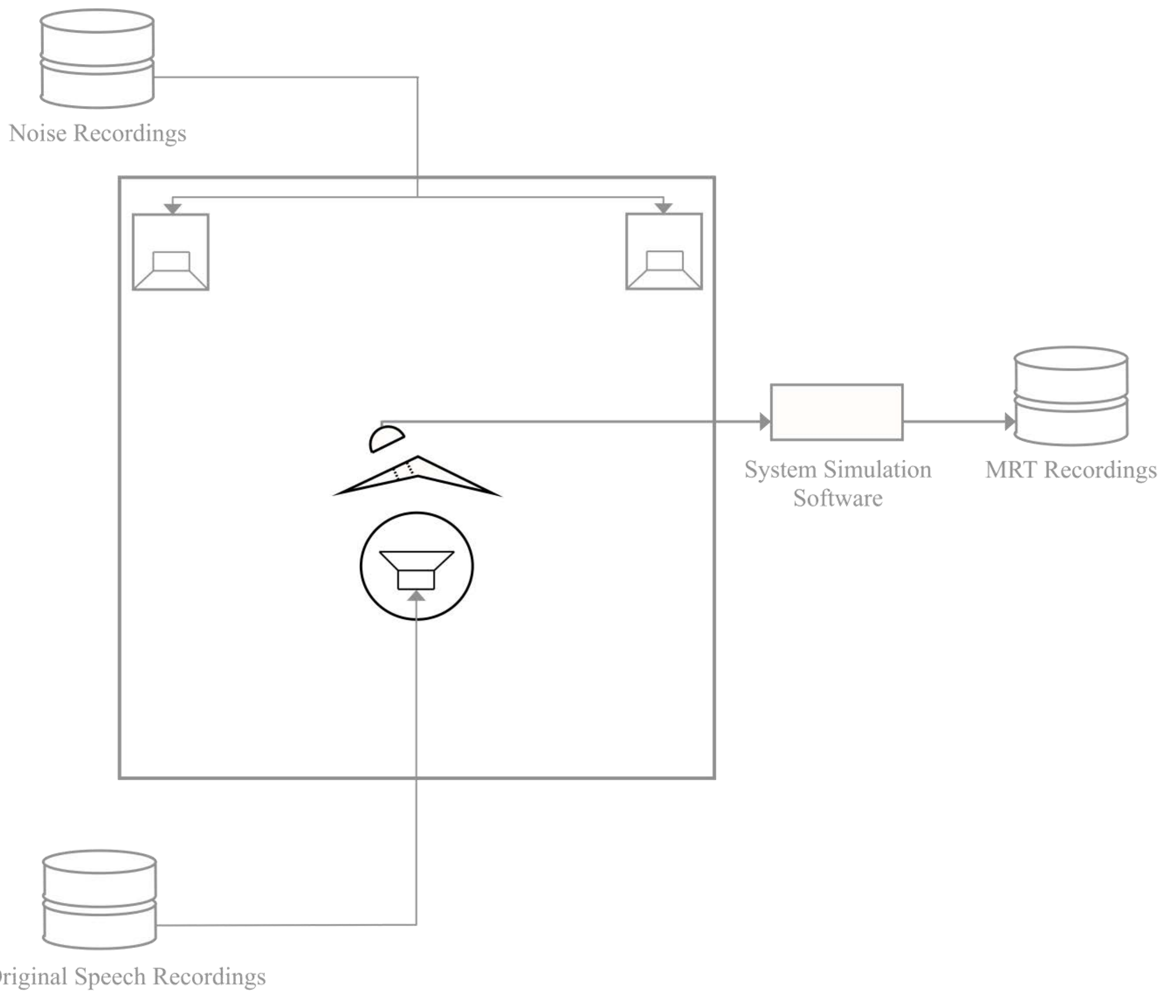


Figure 6. Block diagram for Environments 1 and 2.



,

Figure 7. Block diagram for Environments 3, 5, and 7.

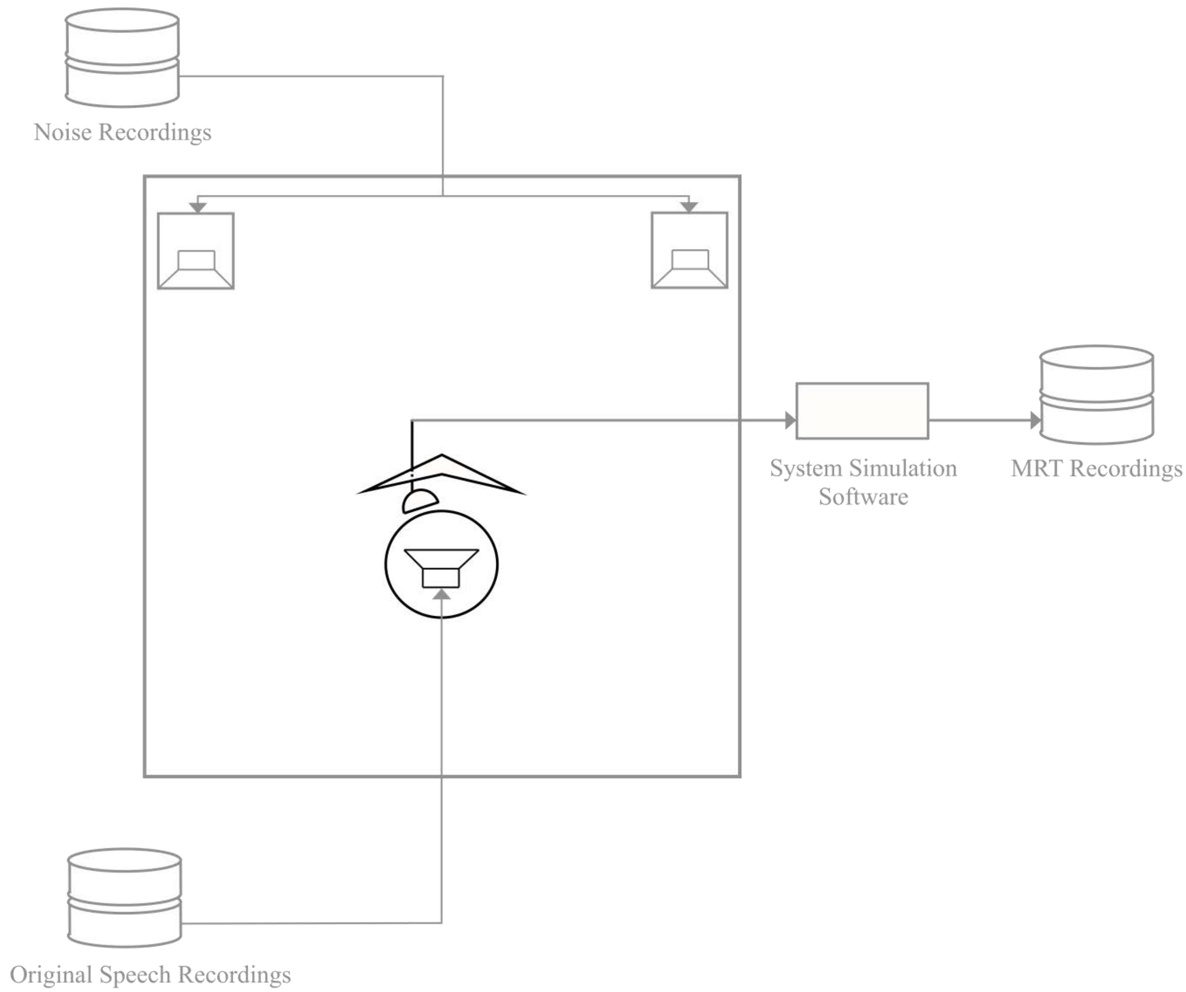


Figure 8. Block diagram for Environments 4 and 6.

4 MODIFIED RHYME TEST

4.1 Listening Lab

We conducted the MRT in a sound-isolated room with inside dimensions 305 cm long, 274 cm wide and 213 cm high (approximately 10 by 9 by 7 feet). The floor is carpeted and all of the walls and the ceiling are covered with sound absorbing materials. Under normal conditions, with no sounds generated inside the room and typical office type noise outside the room, the noise level inside the room was measured at 21.9 dBA with a Brüel and Kjær Type 2250 sound level meter.

This room was set up to meet the standards set forth in sections 8.10.4.11–8.10.4.14 of [7]. These sections pertain to creating a uniform sound field of pink noise, extending from 400 to 4000 Hz. A diagram of the set up can be seen in Figure 9. The loudspeaker carrying the speech signal sat on a table and was placed equidistant from the room side walls, on the edge of the table nearest the subject. The listening position was also equidistant from the room side walls, and 150 cm away from the speech loudspeaker. The two loudspeakers (multiple loudspeakers are allowed by 8.10.4.14) on either side of the table were used to produce the pink noise. Those loudspeakers were pointed toward the back of the room, and were not pointed directly at the subject (8.10.4.12 specifies that the noise field loudspeaker forward axis be oriented away from the position of the listening subject.) The combination of using two loudspeakers to produce the pink noise and the distance from the loudspeakers to the listening position created a quasi-uniform field of sound, thus satisfying 8.10.4.13.

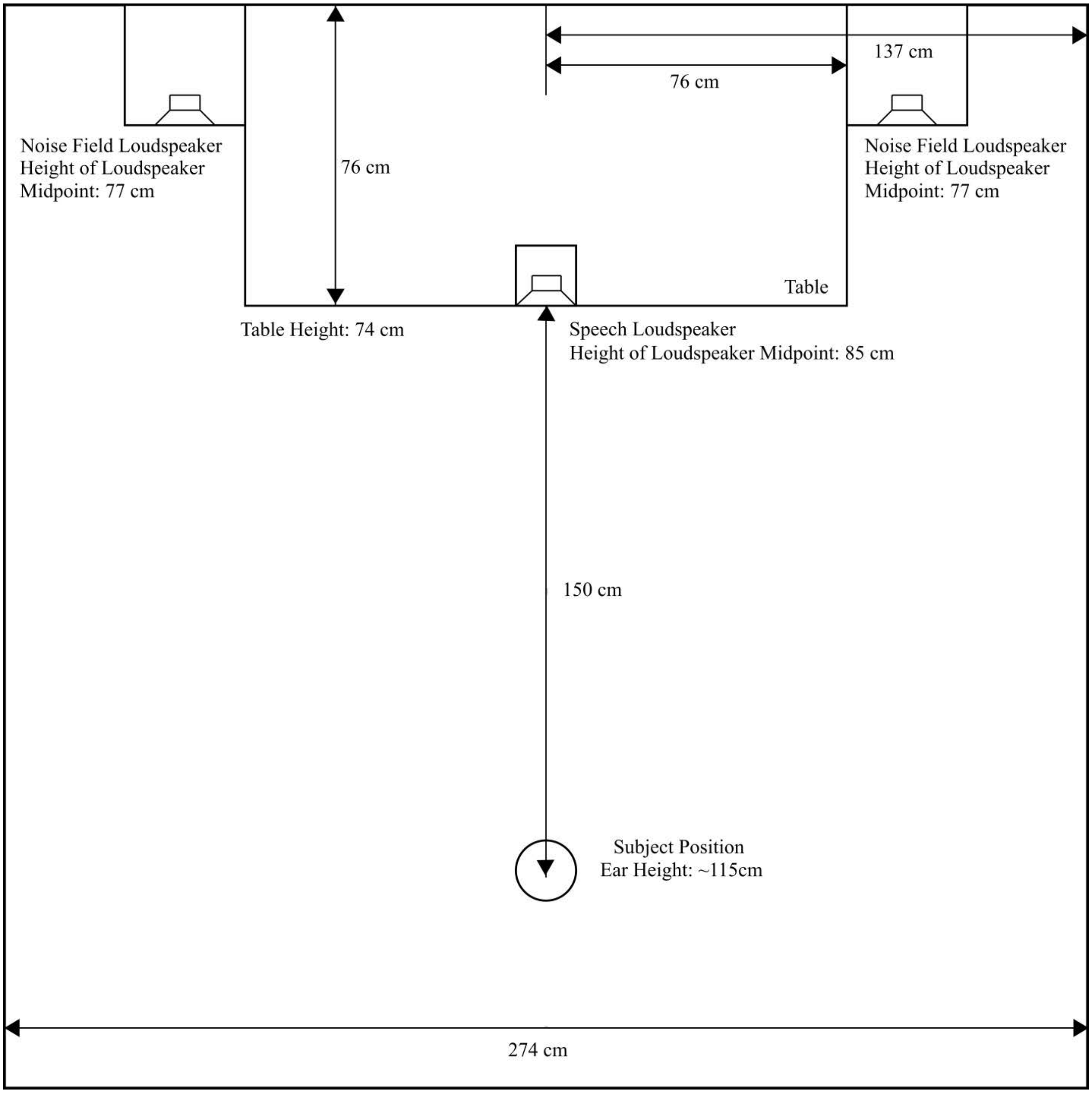


Figure 9. Layout of listening lab for MRT.

In order to generate the field of pink noise in the sound-isolated room as specified in 8.10.4.11, we played a CD recording of the pink noise used in a previous MRT [1] conducted in the same room. From the CD player, the signal passed through a mixer to an equalizer, which then sent the signal to a power amplifier that powered two Electro-Voice Sentry 100A studio monitor loudspeakers in the room. Section 8.10.4.11 of [7] specifies "'pink' noise with a tolerance of 6 dB per octave band from 400 to 4000 Hz." We used a Gold Line DSP30RM real-time spectrum analyzer with one-third octave resolution to analyze the pink noise spectrum. Using information from the spectrum analyzer, the equalizer was adjusted to achieve approximately equal power in each third-octave band.

After equalization was performed with the real-time spectrum analyzer, the noise field was recorded for further, off-line verification. The frequency range from 400 to 4000 Hz covers eleven third-octave bands or three and two-thirds octaves. This allows for multiple interpretations of the "6 dB per octave band" constraint. We chose to analyze two different sets of octave bands, with three bands in each set. Octave bands starting at 400, 800, and 1600 Hz will be considered octave band set A (OBSA), and octave bands starting at 630, 1250, and 2500 Hz will be considered octave band set B (OBSB). Each of these sets is as large as possible without including frequencies outside the range of interest.

The off-line verification started with a third- octave analysis performed in MATLAB®. Energy in OBSA and OBSB were calculated by summing the powers in the corresponding third-octave bands. The results for OBSA and OBSB are shown in Tables 1 and 2 respectively. Since only relative powers are of interest, we arbitrarily normalized our first result to 0 dB.

Table 1. Relative power in three octaves for octave band set A.

Octave Band	**Center Frequencies of Third-Octave Bands Used**	**Relative Power**
1	400, 500, 630 Hz	0.0 dB
2	800, 1000, 1250 Hz	+0.5 dB
3	1600, 2000, 2500 Hz	-0.3 dB

Table 2. Relative power for three octaves for octave band set B.

Octave Band	**Center Frequencies of Third-Octave Bands Used**	**Relative Power**
1	630, 800, 1000 Hz	-0.9 dB
2	1250, 1600, 2000 Hz	+1.6 dB
3	2500, 3150, 4000 Hz	-3.7 dB

These tables make it clear that regardless how one choses to define octaves, the octave-to-octave power variation never exceeded 5.3 dB, thus conforming with the 6 dB constraint on spectral variation. The Brüel and Kjær Type 2250 sound level meter was used to set the presentation level of the noise at the subject's head position, with no subject present. We selected a target presentation level of 65 dBA and the measured result was 64.9 dBA.

The MRT speech recordings originated in MATLAB running on a PC, propagated through an external (connected by FireWire®) audio interface for digital-to-analog conversion, to a mixer, and then finally to the powered studio monitor, Model V4 from KRK Systems LLC. When the signal path was active but no signal propagated to the speakers, the noise level in the chamber fell to 21.9 dBA, measured with the Brüel and Kjær Type 2250 sound level meter. When MRT recordings were played, the average speech level measured 84.2 dBA.

4.2 An MRT Trial

In the MRT a subject heard a carrier sentence (e.g., “Please select the word bed”) and then performed that task. In our implementation the task was performed on a graphical user interface (GUI) displayed on an iPod Touch. The screen size was approximately 5.1 cm wide by 7.6 cm high. An image of the GUI presented on the screen is shown in Figure 10. The subject performed the task by pressing the appropriate button, using a finger or a stylus. There were six words to choose from and the order in which they appeared (top to bottom) was randomly selected by software before the GUI for the trial was displayed. Once a word was selected, the “Play” button appeared, allowing the subject to start the next trial. MRT recordings were only played one time. This whole process describes one MRT trial.

Figure 10. Screenshot of MRT graphical user interface.

If the recording had very high intelligibility, the word “bed” would have been easy to distinguish and the vast majority of the trials would have led to the selection of the correct answer (e.g., the “bed” button would have been selected). If the recording was one with very low intelligibility, subjects may have heard “led,” “fed,” “red,” “wed,” or “shed” instead of “bed.” Alternatively, the subjects may simply have guessed which of the six words was presented. In this case the vast majority of the trials would have led to the selection of an incorrect answer (e.g., “led,” “fed,” “red,” “wed,” or “shed”).

MRT trials were performed repeatedly on each condition (a total of 950 to 1200 trials depending on the condition) and each trial was classified as a success or a failure. This provides the raw data for further statistical analysis.

4.3 MRT Structure

The MRT consisted of a practice session and seven test sessions. The practice session contained groups of five consecutive trials from randomly selected conditions. There were 24 such groups, resulting in a total of 120 trials. The practice session allowed the subjects to familiarize themselves with the MRT process and to resolve implementation issues before the actual test began. It also exposed subjects to the range of recordings that would be encountered in the actual test. In addition, the practice session allowed the test administrator to confirm proper operation of the all equipment involved in the test before any actual data was acquired. The data from the practice trials was not used in subsequent statistical analysis.

Actual sessions contained groups of 50 consecutive trials called blocks. Each block used a single talker and a single condition (e.g., Female 1, System 4, Environment 7.) Each block contained a trial based on each of the 50 MRT word sets. The word set presentation order within each block was random. The key word that was presented in the carrier sentence from each word set was also chosen at random.

Here is an example. Suppose block 1 began with words sets 14, 3, 27, 8, … and block 2 began with word sets 29, 4, 14, 27, … . The sequence of correct answers for block 1 was “went,” “hold,” “pat,” “lane,” …. The sequence of correct answers for block 2 was “kit,” “must,” “dent,” “pad” … Note that word set 14 was used for trial 1 of block 1 and for trial 3 of block 2. That word set contains the words “went,” “sent,” “bent,” “dent,” “tent,” and “rent.” The first word (“went”) was the correct choice when word set 14 was presented in block 1, but the fourth word (“dent”) was the correct choice when word set 14 was presented in block 2. Similarly, word set 27 appeared in both blocks, but the correct choice was different (“pat” vs. “pad).

These processes of randomization prevented subjects from learning correct answers or patterns of answers as the MRT progressed. By preventing any repeatable patterns in the presentation of trials, we can be certain that the subjects’ answers were based solely on their ability to understand the key word in the current trial.

After each block of 50 trials with one condition and talker, a different condition and a different opposite-sex talker were randomly selected. Thus the gender of the talker alternated between female and male with each group of 50 trials. Six groups of 50 trials comprised a session of 300 trials. Each subject completed seven such sessions for a total of 42 groups of 50 trials and a grand total of 2100 trials. We used random presentation order of conditions in order to reduce any possible presentation order effects.

4.4 Test Subjects and Procedure

In light of the context of the MRT, we chose to use public safety practitioners as test subjects. Practitioners have experience communicating in noisy environments, through masks, and over radio links, thus making their MRT results especially relevant to the question at hand.

We recruited the participation of 15 subjects from the public safety community. Eight subjects were firemen, two identified as law enforcement and one identified as a federal responder (four did not respond to the survey question). Eleven subjects reported service lengths ranging from 5 to 32 years with an average of 16.3 years. Subjects were from various locations spanning the U.S.

Fourteen of the subjects were male and one was female. Of the 12 subjects that reported their age, four were between 20 and 29 years old, two were between 30 and 39 years old, four were between 40 and 49 years old, and two were between 50 and 59 years old. According to standard protocol for experiments with human subjects, each subject read and signed a statement of informed consent.

For each subject we measured pure-tone hearing thresholds at 500, 750, 1000, 1500, 2000, and 3000 Hz for each ear, in a quiet booth, as specified in [7] (which then refers to [8]). Ten subjects were found to have thresholds that follow the definition of "audiometrically normal" given in Section 5.3 of [8]. This definition specifies "having hearing threshold levels that are no higher than +20 dB and no lower than -10 dB at any audiometric test frequency, …" For the remainder of this report, this particular definition is implied each time we use the term "audiometrically normal."

Five subjects had thresholds higher than 20 dB HL at one or more of the test frequencies in at least one ear, preventing them from being classified as "audiometrically normal." All fifteen subjects participated in the MRT—most had traveled from out-of-state in order to participate. The hearing threshold issue and its relationship to MRT results are treated in Section 5.2.

Next we seated the subject in a comfortable chair located at the listening position shown in Figure 9. The subject then received verbal instructions regarding the basic MRT task, and the schedule and overall length of the MRT. Any procedural questions asked by the subjects were answered. However, in the interest of avoiding any potential biases, questions regarding the motivation, content, or expected outcomes of the MRT were deferred until after the completion of testing.

The test commenced with the practice session, as described in Section 4.3. After the practice session the test administrator checked to see if the subject had any questions, or had encountered any difficulties. After resolving any issues, the test moved to Session 1. The administrator checked on subjects after each session. Breaks between sessions were offered, but for the most part subjects preferred to continue and simply move on to the next session. Approximately six of the subjects elected to take a break after Session 3 or 4 and the typical break duration was ten to fifteen minutes. Most test sessions lasted between 20 and 30 minutes; this corresponds to an average of 4 to 6 seconds per trial.

5 TEST RESULTS AND DISCUSSION

5.1 Analysis of MRT Trials

In the MRT described here, 15 subjects each completed 2100 trials for a grand total of 31,500 trials. The MRT achieved exact balance of talker gender: 15,750 trials with female talkers and the same number with male talkers. The MRT achieved approximate balance of talkers within gender. For female talkers, Talker 1was heard in 51.1% of the trials and Talker 2 was heard in the remaining 48.9%. For male talkers, Talker 1 was heard in 49.2% of the trials, and Talker 2 in the remaining 50.8%.

On a per-condition basis, the gender balance ranged from approximate (57.1% female) to exact (50.0% female). An exact balance of the trials across the 28 conditions would require that 3.6% of the trials be allotted to each condition. The MRT achieved approximate balance in this respect: the various conditions received between 3.0% and 3.8% of the trials. Of the 31,500 trials performed, 4,600 trials were associated with the four invalid conditions described in Section 3.2 and thus are not included in the analysis that follows. This leaves 26,900 trials for analysis.

We view the MRT as set of repeated trials. Each trial can be classified as a success (the proper key word was selected on the GUI) or a failure (a word other than the proper key word was selected on the GUI). Since each trial results in success or failure, Bernoulli trials and the underlying binomial distribution provide a model for these trials [13]. In a Bernoulli trial there are exactly two possible outcomes and these are generically labeled as "success" and "failure." The probability of success is specified by the variable p.

For any group of N trials bearing x successes, we can find a maximum likelihood estimate $\hat{p}$ of the underlying parameter p. It turns out that this statistically rigorous estimate aligns well with intuition:

$$\hat{p} = \frac{x}{N}. \tag{1}$$

That is, the estimated probability of success in the underlying Bernoulli model is simply the fraction of successes observed. This estimated probability of success provides the basis for reporting intelligibility. Because the MRT offered six word choices, the worst possible probability of success is $\frac{1}{6} = 0.167$. In other words, even with the speech signal turned off (clearly a case of zero intelligibility), one could select the correct word 16.7% of the time, simply by selecting one of the six word options at random. Thus [8] specifies a transformation that maps $\hat{p}$ to intelligibility:

$$Intelligibility = \frac{6}{5}\left(\hat{p} - \frac{1}{6}\right). \tag{2}$$

This relationship maps $\hat{p} = \frac{1}{6}$ (the success rate for guessing) to $Intelligibility = 0$. It also maps $\hat{p} = 1$ to $Intelligibility = 1$, as desired.

For each of the 24 conditions, we combine all trials from only the ten “audiometrically normal” subjects to produce a single intelligibility value. These results are displayed graphically in Figure 11. In this figure seven clusters of bars represent the seven environments. Within each cluster, the bars indicate the MRT intelligibility results for Analog, P25, AMR12.2 and AMR7.4, in that order (where available). (The MRT intelligibility results for the four erroneous conditions are 0.52, 0.48, 0.52, and 0.41.)

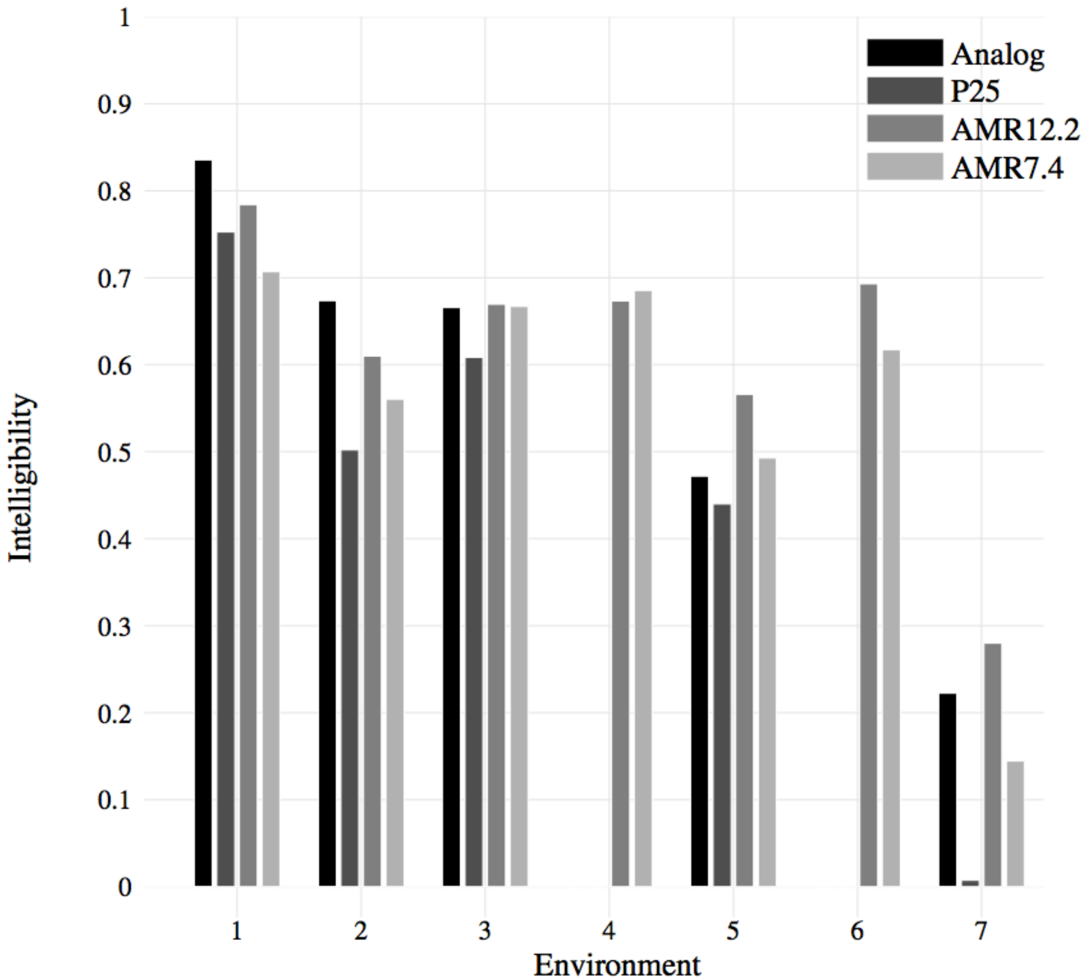


Figure 11. Intelligibility results for 24 conditions and ten “audiometrically normal” subjects.

Before drawing conclusions from Figure 11, we must acknowledge that there is some uncertainty in those results. This uncertainty is inherent in any measurement or testing process and can be quantified and accounted for through various statistical techniques. The chi-squared statistic, denoted χ^2, can be used to test the hypothesis that two Bernoulli processes have the same value of p [14][15]. Thus it can be used to compare MRT results for two systems to see if they are the same or not.

More formally, we construct the null hypothesis: System A and System B have the same MRT success rates. Next we construct a χ^2 statistic that compares the success rates of System A and System B. If that statistic exceeds a critical value (determined by our desired significance level) then we can reject the null hypothesis. Suppose observation of System A yields x_A successes in N trials. We use System B as a reference system. System B has $\hat{p} = p_B$ so we would expect it to produce $x_B = p_B N$ successes and $N - x_B$ failures in N trials. The χ^2 statistic allows us to compare the observed values from System A with the expected values associated with the reference system, System B:

$$\chi^2 = \frac{(x_A - x_B)^2}{x_B} + \frac{((N - x_A) - (N - x_B))^2}{(N - x_B)}. \tag{3}$$

If

$$3.841 \le \chi^2, \tag{4}$$

we can reject the hypothesis that System A (the observed system) and System B (the reference system that sets expectations) have the same MRT success rates (and thus the same speech intelligibilities) at the 95% level of significance.

We are interested in the relative performance of the four systems in each environment. In five of the environments six distinct comparisons are possible. Four of these are of primary interest—they compare AMR with the two reference systems: AMR12.2 vs. Analog, AMR12.2 vs. P25, AMR7.4 vs. Analog, and AMR7.4 vs. P25. In addition, we can compare the two forms of AMR: AMR12.2 vs. AMR7.4. Finally, while it is not a goal of this work, one can also compare the two reference systems: Analog vs. P25.

The MRT results are such that in the majority of the cases, comparison of two systems in a fixed environment results in rejection of the hypothesis that two systems have the same intelligibility ($3.841 \le \chi^2$). Thus in Figures 12 and 13 we mark the cases where the hypothesis is not rejected. If the hypothesis is not rejected, we say the systems are statistically equivalent at the 95% level. Toward that end, Figures 12 and 13 use a solid horizontal bar to indicate if two or more systems in a given environment are statistically equivalent. In Environments 2, 6, and 7, none of the system comparisons show statistical equivalence at the 95% level.

These results are also presented in tabular form in Table 3. This table steps through all possible system comparisons for each of the seven environments and details the outcome of all those comparisons. Further discussion of these results begins in 5.3.

Note that (3) requires one to choose a reference system and an observed system. This gives two possible results for each pair of systems that are compared. For example, when comparing AMR12.2 with AMR7.4 one could pick AMR12.2 to be the reference, or one could pick AMR7.4 to be the reference. We have done all comparisons both ways and while the resulting values of χ^2 do change, the results of the test $3.841 \le \chi^2$ do not change. In other words, all results in Figure 12, Figure 13, and Table 3 are invariant to the assignment of "reference system" and "observed system."

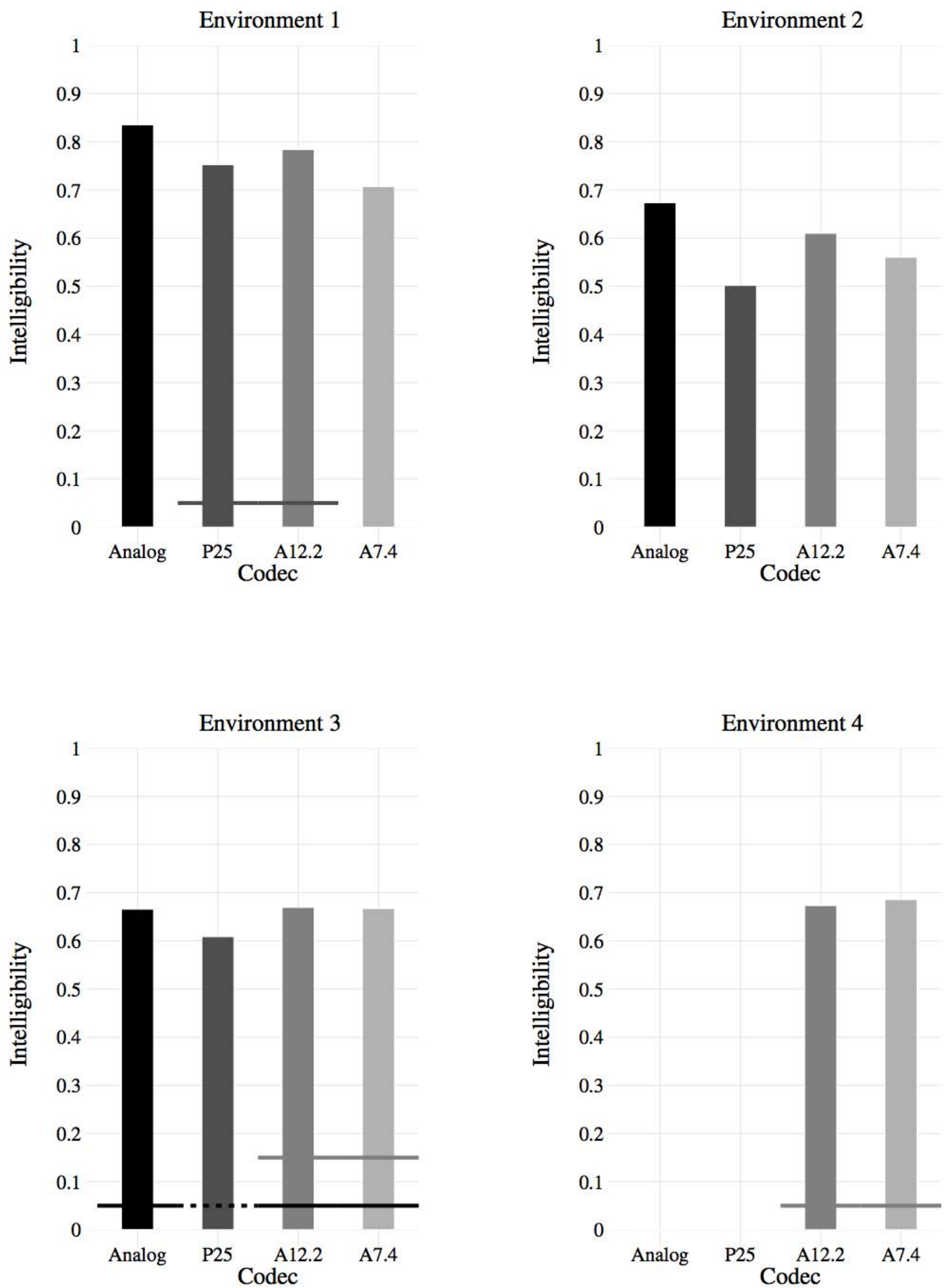


Figure 12. Intelligibility results for ten "audiometrically normal" subjects, Environments 1–4, showing statistical similarities.

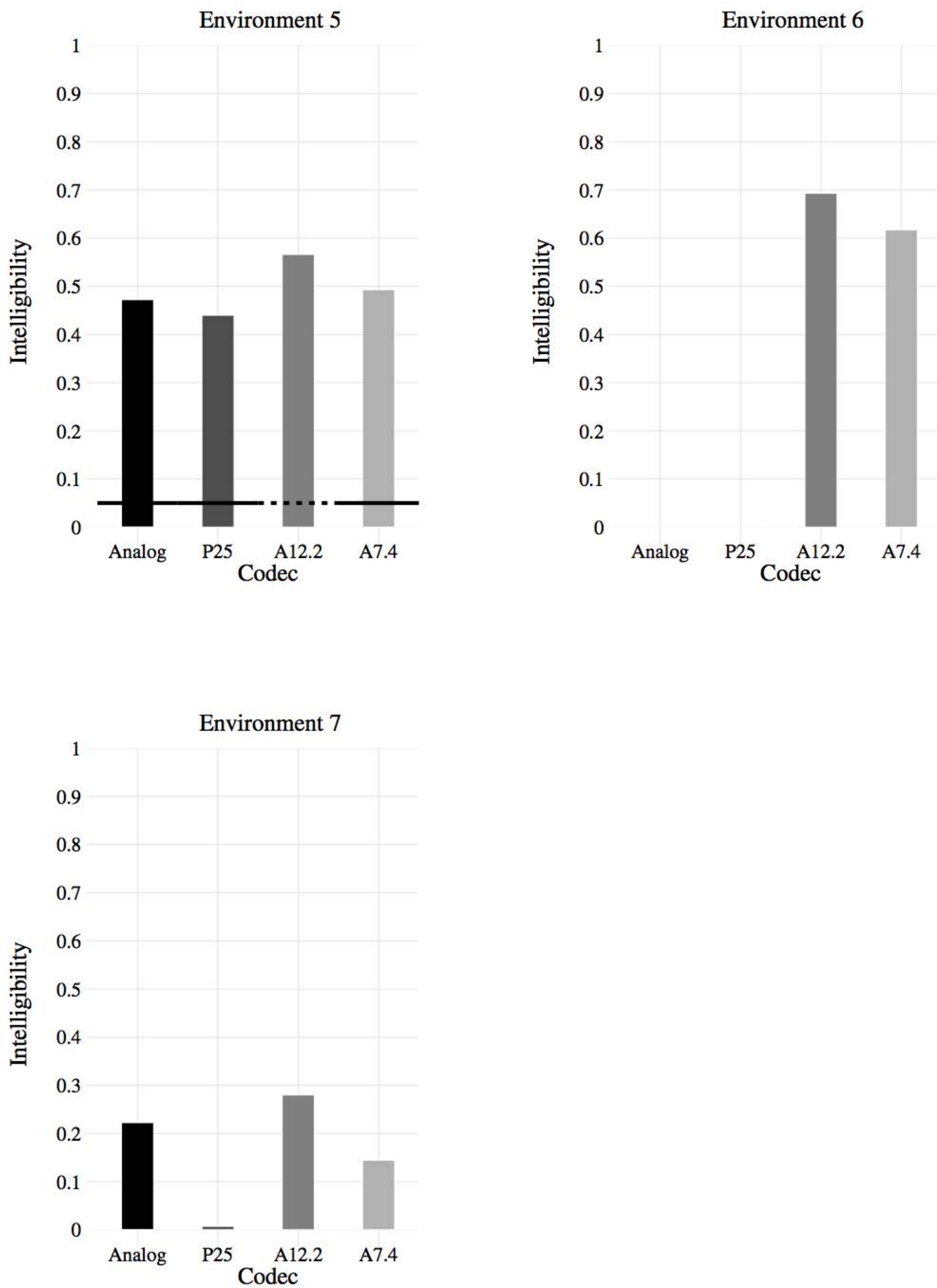


Figure 13. Intelligibility results for ten “audiometrically normal” subjects, Environments 5–7, showing statistical similarities.

Table 3. Significant differences in intelligibility between systems for each environment.

Environment	10 Subjects with "Audiometrically Normal" Pure Tone Thresholds	All 15 Subjects
1 (No noise added)	AMR12.2 < Analog AMR12.2 ≡ P25 AMR12.2 > AMR7.4 AMR7.4 < Analog AMR7.4 < P25 Analog > P25	AMR12.2 < Analog AMR12.2 ≡ P25 AMR12.2 > AMR7.4 AMR7.4 < Analog AMR7.4 < P25 Analog > P25
2 (Nightclub noise added)	AMR12.2 < Analog AMR12.2 > P25 AMR12.2 > AMR7.4 AMR7.4 < Analog AMR7.4 > P25 Analog > P25	AMR12.2 < Analog AMR12.2 > P25 **AMR12.2 ≡ AMR7.4** AMR7.4 < Analog AMR7.4 > P25 Analog > P25
3 (No noise added, mask, microphone at vox port)	AMR12.2 ≡ Analog AMR12.2 > P25 AMR12.2 ≡ AMR7.4 AMR7.4 ≡ Analog AMR7.4 > P25 Analog > P25	AMR12.2 ≡ Analog AMR12.2 > P25 AMR12.2 ≡ AMR7.4 AMR7.4 ≡ Analog AMR7.4 > P25 Analog > P25
4 (No noise added, mask, microphone in mask)	AMR12.2 ≡ AMR7.4	AMR12.2 ≡ AMR7.4
5 (PASS alarm, mask, microphone at vox port)	AMR12.2 > Analog AMR12.2 > P25 AMR12.2 > AMR7.4 AMR7.4 ≡ Analog AMR7.4 > P25 Analog ≡ P25	**AMR12.2 ≡ Analog** AMR12.2 > P25 AMR12.2 > AMR7.4 AMR7.4 ≡ Analog **AMR7.4 ≡ P25** Analog ≡ P25
6 (PASS alarm, mask, microphone in mask)	AMR12.2 > AMR7.4	AMR12.2 > AMR7.4
7 (Chainsaw noise added, mask, microphone at vox port)	AMR12.2 > Analog AMR12.2 > P25 AMR12.2 > AMR7.4 AMR7.4 < Analog AMR7.4 > P25 Analog > P25	AMR12.2 > Analog AMR12.2 > P25 AMR12.2 > AMR7.4 AMR7.4 < Analog AMR7.4 > P25 Analog > P25

When the hypothesis that the two systems have the same intelligibility has been rejected at the 95% level, the symbol "<" or ">" indicates which system has greater intelligibility. Otherwise "≡" indicates statistical equivalence of intelligibility. Bold font indicates a result changed when the subject group is increased from ten to fifteen.

5.2 Elevated Hearing Thresholds

In addition to the ten subjects with "audiometrically normal" pure tone thresholds, we also have MRT results for five subjects with elevated hearing thresholds. One could argue that from a purely experimental perspective, the fairest and most accurate results will be found through maximum control of variables within the experiment. Maximum control implies excluding subjects with elevated thresholds. This is the action specified in [8] and is also the basis for Figure 11.

On the other hand, the recruited group of fifteen subjects forms a small (and not completely random) sample of the public safety practitioner community. It is expected that the entire population of practitioners will contain persons with normal thresholds as well as persons with elevated thresholds, so it is not surprising that this sample contains both as well. Thus from a more practical perspective, one might argue that the more realistic and relevant results would come from including both subjects with normal thresholds and subjects with elevated thresholds. One cannot, however, argue that the mixture of pure tone hearing thresholds present in this sample of 15 subjects accurately reflects the mixture present in the entire practitioner population. That goal could only be achieved through a much larger, completely random sample.

It is likely that some readers will find themselves more aligned with one set of the arguments above while others will be swayed by the other set of arguments. It might boil down to those seeking the most rigorous results versus those seeking the most relevant results. To address this situation we provide results based on the entire group of fifteen subjects in Figure 14, in addition to the result for the group of ten "audiometrically normal" subjects already presented in Figure 11. Likewise, Table 3 contains a column of results based on the entire group of fifteen subjects, as well as a column of results base on the group of ten. Figures 15 and 16 together represent the same data shown in Figure 14 and use a solid horizontal line to indicate two or more systems in a given environment that are statistically equivalent.

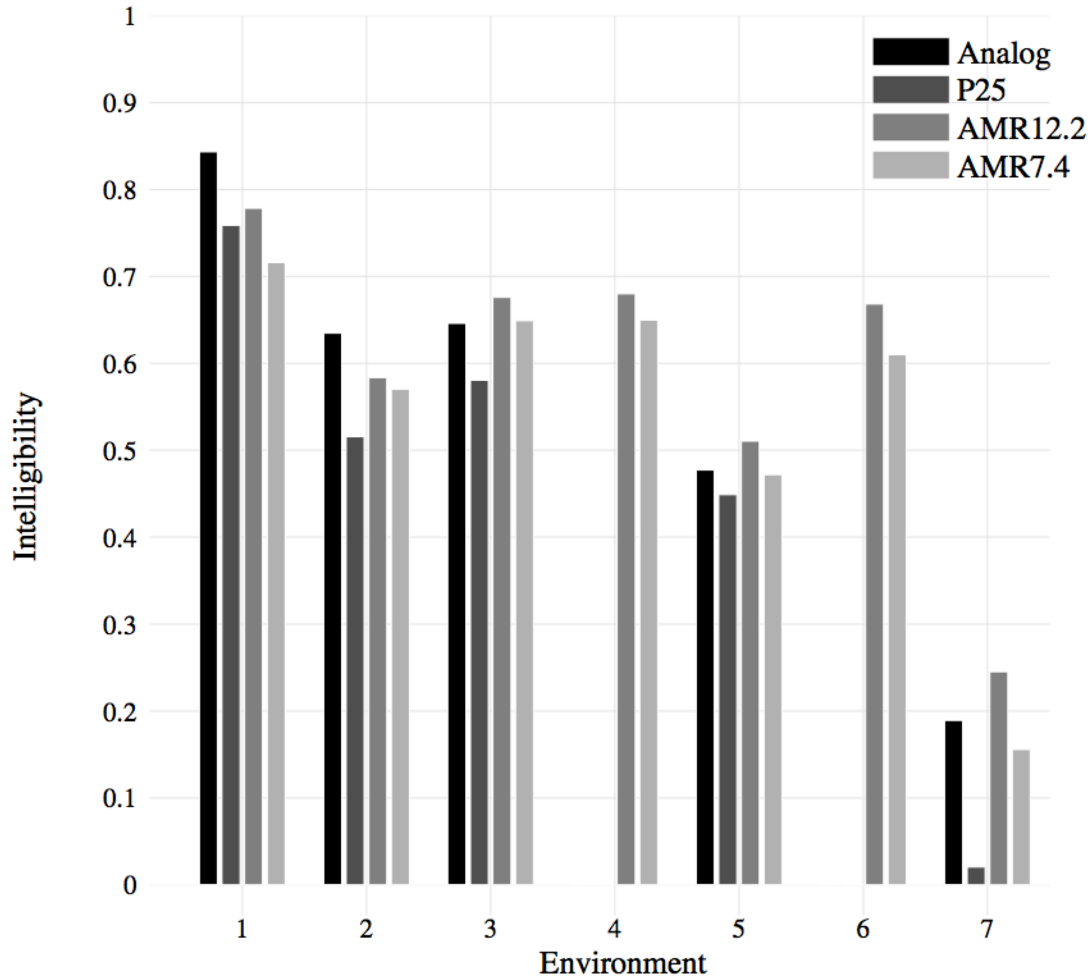


Figure 14. Intelligibility results for 24 conditions and all 15 subjects.

The differences between the two sets of results are minor. When we compare two systems from a given environment, the two sets of results lead to the same conclusions for 29 of the 32 different comparisons. The remaining three cases are indicated in bold font in Table 3.

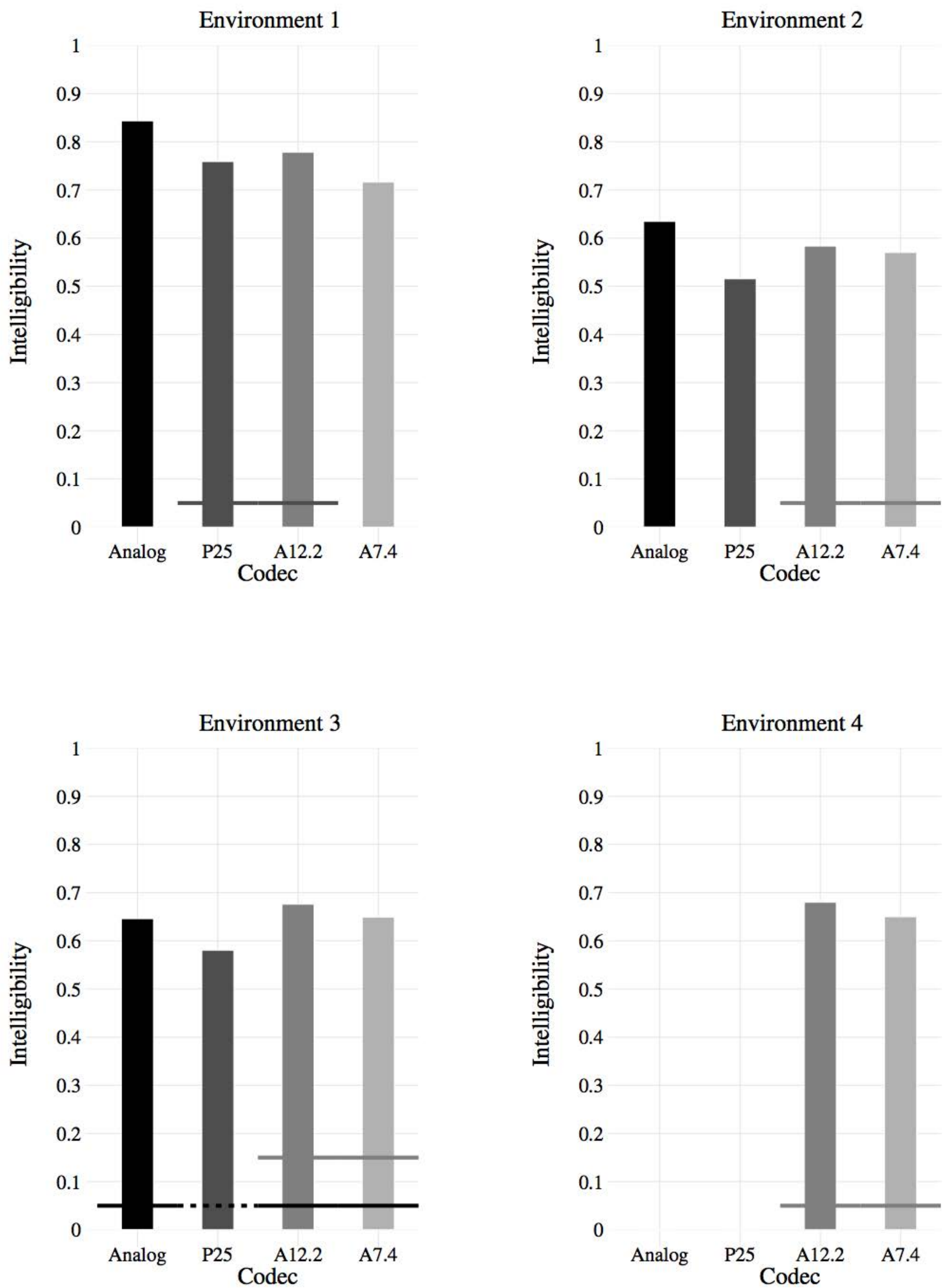


Figure 15. Intelligibility results for all 15 subjects, Environments 1–4, showing statistical similarities.

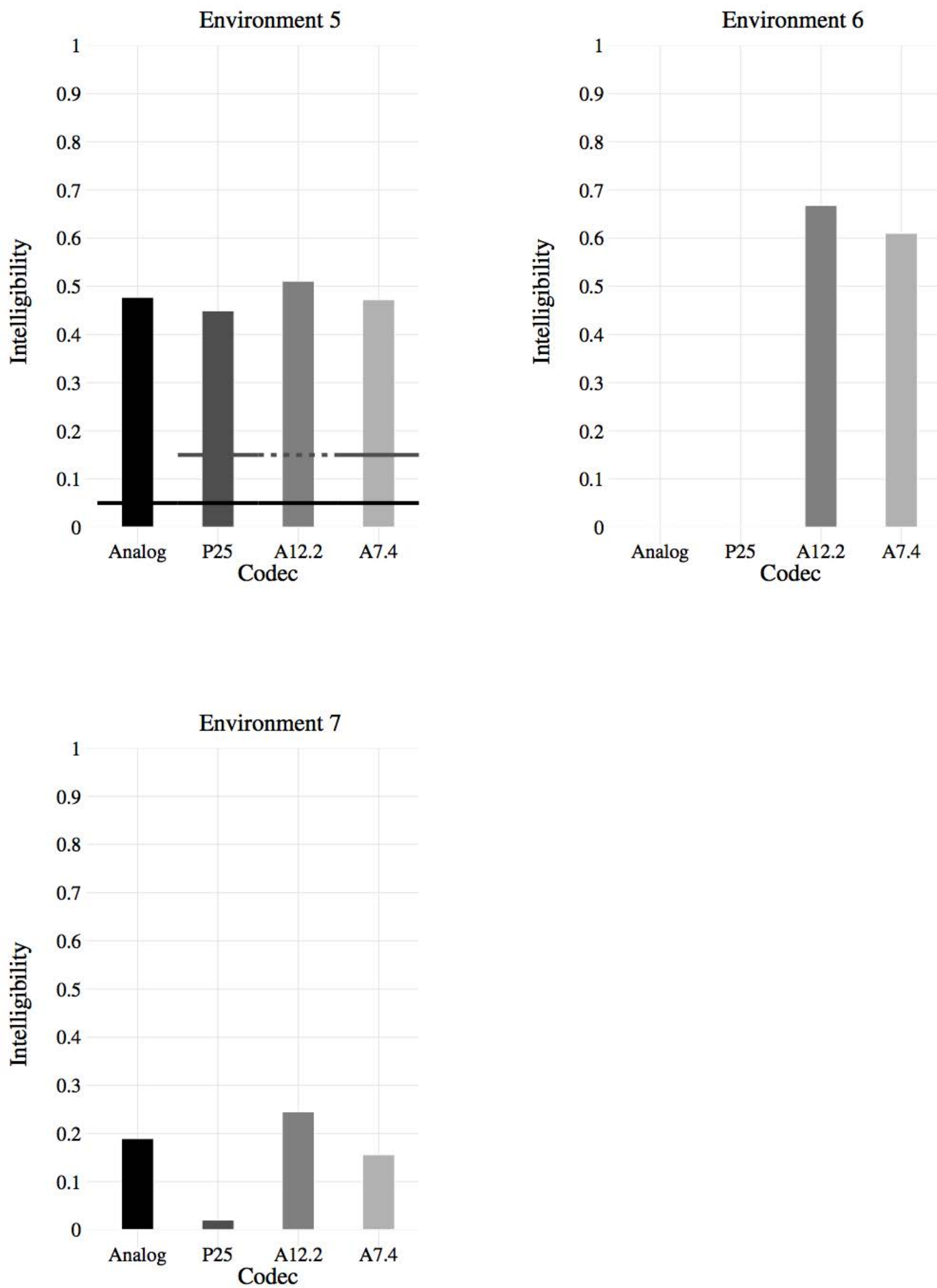


Figure 16. Intelligibility results for all 15 subjects, Environments 5–7, showing statistical similarities

We also offer a more detailed look at the five subjects displaying elevated pure-tone hearing thresholds. In each of the five cases, the elevated threshold occurred at a single test frequency. Thus we can easily summarize these elevated thresholds for each subject in one line:

- 25 dB HL (one ear) at 2 kHz (sorted subject number 4)
- 30 dB HL (one ear) at 3 kHz (sorted subject number 12)
- 25 and 30 dB HL (two ears) at 3 kHz (sorted subject number 5)
- 35 and 40 dB HL (two ears) at 3 kHz (sorted subject number 6)
- 55 dB HL (one ear) at 3 kHz (sorted subject number 1)

Each description above ends with a sorted subject number. These numbers refer a position on to the horizontal axis of Figure 17. This figure shows the sorted overall (using 24 conditions) intelligibility values for each of the 15 subjects, along with a 95% confidence interval for each one. The figure reveals a fairly smooth distribution, punctuated by one or two outliers at the low-intelligibility end, and one outlier at the high-intelligibility end of the distribution.

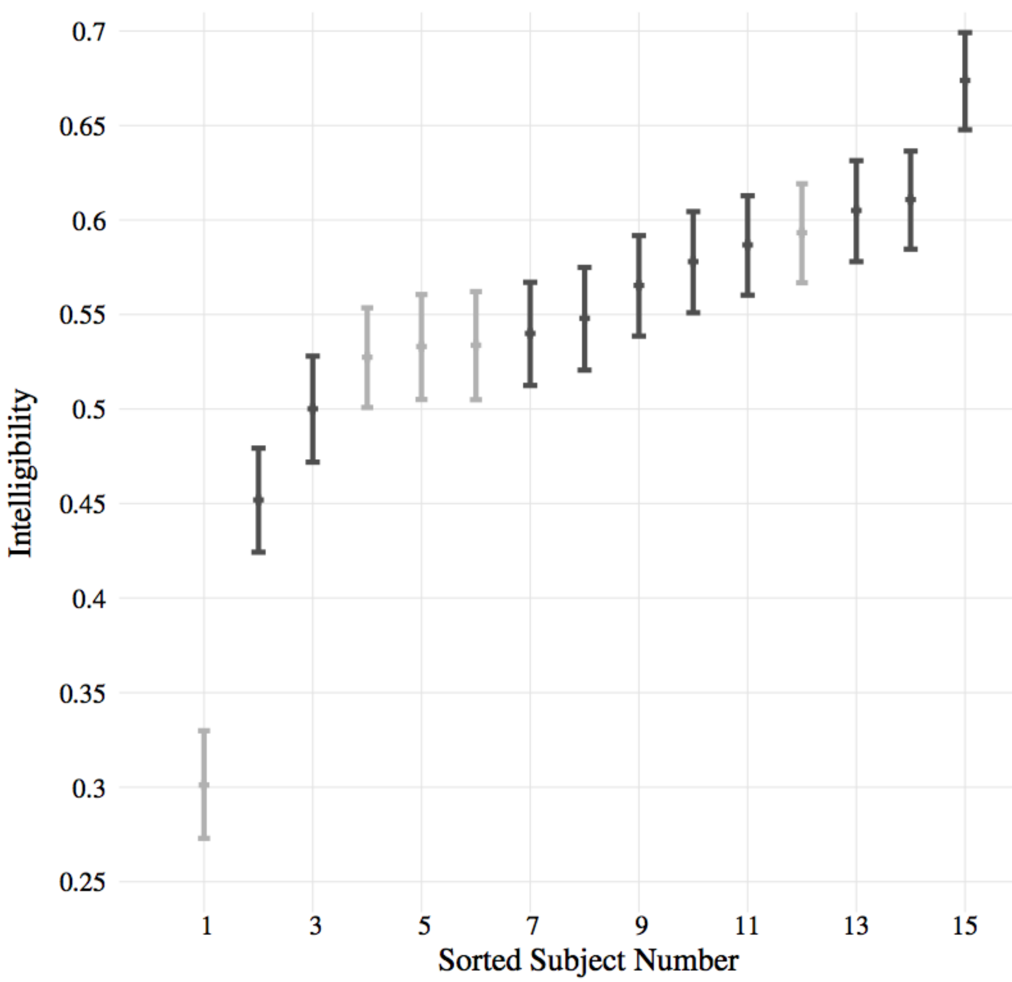


Figure 17. Grand mean intelligibility per subject, sorted.

Each subject result is based on a large number (between 1650 and 1850) of MRT trials that are approximately balanced with respect to talker, system, and environment. Thus we expect that the relative position of a subject in Figure 17 may reflect that subject's innate abilities for hearing and detecting rhyming words in the presences of noises and distortion, as well as that subject's level of effort in applying those innate abilities.

In Figure 17, the five subjects with elevated threshold are shown with light gray markers and 95% confidence intervals. The subject with the lowest overall intelligibility value falls outside the main distribution and is one of the five subjects with an elevated threshold. The other four subjects with elevated thresholds appear as part of the main distribution, though three of the four appear in the lower half of the distribution and only one appears in the upper half. We conclude that only one of the five subjects falls outside the main population in terms of overall intelligibility value.

5.3 Comparison of Systems

The acoustic environments in this MRT ranged from benign (no noise, no mask, median intelligibility near 0.77) to extremely difficult (chainsaw noise and mask, median intelligibility near 0.17). As expected, the four different systems measure up differently under these various, disparate environments and we have provided details for all possible comparisons in Figures 11 and 14. Those results are replicated in Table 3.

Recall from Section 1 the motivating question for this work: *How does the speech intelligibility of the AMR speech coder compare with the speech intelligibility of analog and digital reference systems when used in emergency-response environments?* The results obtained allow us to speak to that question and we summarize the key results in the nine statements that follow. These statements are organized into five categories. Each of the nine statements is true at the 95% significance level regardless of which subject group is analyzed.

Regarding speech intelligibility as measured by the MRT described in this report, using a 95% significance level we find:

AMR12.2 relative to Analog

1. AMR12.2 intelligibility is lower than Analog in Environments 1 and 2.

2. AMR12.2 intelligibility is higher than or the same as Analog in Environments 3, 5, and 7.

AMR12.2 relative to P25

3. AMR12.2 intelligibility is the same as P25 in Environment 1.

4. AMR12.2 intelligibility is higher than P25 in Environments 2, 3, 5, and 7.

AMR7.4 relative to Analog

5. AMR7.4 intelligibility is lower than Analog in Environments 1, 2 and 7.

6. AMR7.4 intelligibility is the same as Analog in Environments 3 and 5.

AMR7.4 relative to P25

7. AMR7.4 intelligibility is lower than P25 in Environment 1.

8. AMR7.4 intelligibility is higher than or the same as P25 in Environments 2, 3, 5, and 7.

AMR12.2 relative to AMR7.4

9. AMR12.2 intelligibility is higher than or the same as AMR7.4 in all seven tested environments.

In summary, AMR12.2 speech intelligibility may be above, the same as, or below Analog intelligibility, depending on the environment. AMR12.2 intelligibility is always above or the same as P25 intelligibility. AMR7.4 speech intelligibility may be the same as or below the Analog intelligibility depending on the environment. AMR7.4 intelligibility is above or the same as P25 intelligibility in all environments except Environment 1. These same conclusions follow from analysis of the group of ten “audiometrically normal” MRT subjects (perhaps the more rigorous option) or the entire group of fifteen MRT subjects (perhaps the more realistic option).

6 REFERENCES

[1] D. J. Atkinson and A. A. Catellier, "Intelligibility of Selected Radio Systems in the Presence of Fireground Noise: Test Plan and Results," NTIA Report TR-08-453, June 2008. Available: www.its.bldrdoc.gov.

[2] Telecommunications Industry Association, TIA-102.BABA-1998, "Project 25 Vocoder Description," May 1998. Reaffirmed December, 2008.

[3] Telecommunications Industry Association, TIA-102.BABA, Annex A, "APCO Project 25 Half-Rate Vocoder Addendum," April 2009.

[4] H.R. 3630: "Middle Class Tax Relief and Job Creation Act of 2012," Washington D.C, February 2012. Available: www.gpo.gov.

[5] 3GPP/ETSI, "AMR speech codec; general description, version 10.0.0," Technical Report TS 26 071, Sophia Antipolis Cedex, France, April 2011.

[6] 3GPP/ETSI, "Performance characterization of the adaptive multi-rate (AMR) speech codec," Technical Report TR 126 975, Sophia Antipolis Cedex, January 2009.

[7] NFPA 1981 Standard on Open-Circuit Self-Contained Breathing Apparatus (SCBA) for Emergency Services, 2007 Edition.

[8] ANSI S3.2, American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems, 1989.

[9] L. Beranek, "Criteria for Noise and Vibration in Communities, Buildings, and Vehicles," Noise and Vibration Control, 1st ed., L. Beranek, Ed New York: McGraw-Hill, pp. 564–566, 1971.

[10] Recommendation ITU-T P.56, "Objective measurement of active speech level," Geneva, December 2011.

[11] Recommendation ITU-T G.191, "Software tools for speech and audio coding standardization," March 2012.

[12] Recommendation ITU-T P.58, "Head and Torso Simulator (HATS) for Telephonometry," Geneva, December 2011.

[13] N. Johnson, S. Kotz, and A. Kemp, "*Univariate discrete distributions, second edition*," New York: Wiley, 1992.

[14] D. C. Montgomery, G. C. Runger, and N. F. Hubele "Engineering statistics, fourth edition," Hoboken, NJ: Wiley, 2007.

[15] R. M. Bethea, B. S. Duran, and T. L. Boullion, "Statistical methods for engineers and scientists, second edition," New York: Dekker, 1985.

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO. TR-13-493	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Intelligibility of the Adaptive Multi-Rate Speech Coder in Emergency-Response Environments		5. Publication Date March 2013 6. Performing Organization Code NTIA/ITS.P
7. AUTHOR(S) David J. Atkinson, Stephen D. Voran, and Andrew A. Catellier		9. Project/Task/Work Unit No. 6692000-320
8. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Telecommunication Sciences National Telecommunications & Information Administration U.S. Department of Commerce 325 Broadway Boulder, CO 80305		10. Contract/Grant Number.
11. Sponsoring Organization Name and Address National Telecommunications & Information Administration Herbert C. Hoover Building 14th & Constitution Ave., NW Washington, DC 20230		12. Type of Report and Period Covered Technical Report

14. SUPPLEMENTARY NOTES

15. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

This report describes speech intelligibility testing conducted on the Adaptive Multi-Rate (AMR) speech coder in several different environments simulating emergency response conditions and especially fireground conditions. The intelligibility testing protocol was the Modified Rhyme Test (MRT). Conditions included background noises of various types, as well as a mask associated with a self-contained breathing apparatus. Analog FM radio transmission and Project 25 digital radio transmission were also included in the test as reference points. Test participants were persons employed as first responders in public safety fields. Through statistical analysis of 26,900 MRT trials we are able to draw conclusions on speech intelligibility for AMR speech coding relative to analog and digital radio reference points for each of the five different operating environments.

16. Key Words (Alphabetical order, separated by semicolons)

AMBE; AMR; analog FM; LMR; fireground; MBE; MRT; P25; public safety; SCBA mask; speech coder; speech intelligibility; subjective testing;

17. AVAILABILITY STATEMENT	18. Security Class. (This report)	20. Number of pages
☒ UNLIMITED.	Unclassified	51
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. (This page) Unclassified	21. Price:

NTIA FORMAL PUBLICATION SERIES

NTIA MONOGRAPH (MG)

A scholarly, professionally oriented publication dealing with state-of-the-art research or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA SPECIAL PUBLICATION (SP)

Conference proceedings, bibliographies, selected speeches, course and instructional materials, directories, and major studies mandated by Congress.

NTIA REPORT (TR)

Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities. Subsets of this series include:

NTIA RESTRICTED REPORT (RR)

Contributions that are limited in distribution because of national security classification or Departmental constraints.

NTIA CONTRACTOR REPORT (CR)

Information generated under an NTIA contract or grant, written by the contractor, and considered an important contribution to existing knowledge.

JOINT NTIA/OTHER-AGENCY REPORT (JR)

This report receives both local NTIA and other agency review. Both agencies' logos and report series numbering appear on the cover.

NTIA SOFTWARE & DATA PRODUCTS (SD)

Software such as programs, test data, and sound/video files. This series can be used to transfer technology to U.S. industry.

NTIA HANDBOOK (HB)

Information pertaining to technical procedures, reference and data guides, and formal user's manuals that are expected to be pertinent for a long time.

NTIA TECHNICAL MEMORANDUM (TM)

Technical information typically of less breadth than an NTIA Report. The series includes data, preliminary project results, and information for a specific, limited audience.

For information about NTIA publications, contact the NTIA/ITS Technical Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. (303) 497-3572 or e-mail info@its.bldrdoc.gov.

This report is for sale by the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161,Tel. (800) 553-6847.