NTIA Report 15-520

Speech Codec Intelligibility Testing in Support of Mission-Critical Voice Applications for LTE

Stephen D. Voran
Andrew A. Catellier

report series

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

NTIA Report 15-520

Speech Codec Intelligibility Testing in Support of Mission-Critical Voice Applications for LTE

Stephen D. Voran
Andrew A. Catellier

U.S. DEPARTMENT OF COMMERCE

September 2015

DISCLAIMER

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.

PREFACE

The work described in this report was performed by the Public Safety Communications Research Program (PSCR) on behalf of the Department of Homeland Security (DHS) Science and Technology Directorate. The objective was to quantify the speech intelligibility associated with a range of digital audio coding algorithms in various acoustic noise environments. This report constitutes the final deliverable product for this project. The PSCR is a joint effort of the National Institute for Standards and Technology and the National Telecommunications and Information Administration.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS/ACRONYMS

3GPP	Third Generation Partnership Project
AAC-LD	Advanced Audio Coding – Low Delay
AAC-ELD	Advanced Audio Coding – Enhanced Low Delay
AAC-ELDsbr	Advanced Audio Coding – Enhanced Low Delay with Spectral Band Replication
ABC-MRT	Articulation Band Correlation Modified Rhyme Test
ADPCM	Adaptive Differential Pulse Code Modulation
AES/EBU	Audio Engineering Society/European Broadcasting Union
AFM	Analog FM
AMBE	Advanced Multi-Band Excitation
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
ANSI	American National Standards Institute
b/smp	bits/sample
DHS	Department of Homeland Security
EVS	Enhanced Voice Services
FB	Fullband
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
kb/s	kilobits/second
LMR	Land Mobile Radio
LTE	Long-Term Evolution

MCV	Mission-Critical Voice
MLT	Modulated Lapped Transform
MPEG	Moving Picture Experts Group
MRT	Modified Rhyme Test
NB	Narrowband
NPSTC	National Public Safety Telecommunications Council
P25	Project 25
PCM	Pulse Code Modulation
PSCR	Public Safety Communications Research Program
SNR	Signal-to-Noise Ratio
TIA	Telecommunications Industry Association
USB	Universal Serial Bus
WB	Wideband

EXECUTIVE SUMMARY

This report describes an effort to quantify the speech intelligibility associated with a range of narrowband, wideband, and fullband digital audio coding algorithms in various acoustic noise environments. The work emphasizes the relationship between these intelligibility results and analogous ones for an analog FM land-mobile radio reference.

The report begins with background information and context for the project. It then describes the creation of speech and noise recordings. These recordings include 54 different noise environments and 1200 different test sentences that follow the modified rhyme test (MRT) paradigm. The recordings were processed by 83 different narrowband, wideband, and fullband digital audio codec modes and by an analog FM reference as well.

The resulting recordings were then processed by an objective estimator of speech intelligibility to gain preliminary insights into the intelligibility of the various codec modes relative to that of analog FM. These results allowed for the design of a practically sized MRT involving 6 challenging yet relevant noise environments and 28 codec modes (a total of 168 conditions, 24 of which are considered reference conditions that anchor the results).

The MRT produced 432 trials for each of the 168 conditions under test (72,576 total trials). This report provides full details of the MRT design, implementation, and administration. Thirty-six subjects from the public safety community participated in the MRT and summary demographics are provided.

A statistical analysis of the MRT results shows that 55 of the 144 non-reference conditions tested yielded intelligibility equivalent to that of the analog FM reference. Thirty-four of the non-reference conditions yielded intelligibility lower than that of analog FM and 55 of the non-reference conditions yielded intelligibility higher than that of analog FM.

MRT results show that intelligibility depends strongly on noise environments. In a quiet environment 21 codec modes match or exceed the intelligibility of the analog FM reference. And 19 codec modes produce intelligibility no lower than analog FM in at least five of the six noise environments. This list includes three narrowband, eleven wideband, and five fullband codec modes with data rates ranging from 6.6 to 48 kbps. But when all six noise environments are considered, only six codec modes consistently produce intelligibility no lower than analog FM. The data rates for these six codec modes range from 16.4 to 32 kb/s.

We expect that the detailed results contained herein can inform some of the design and provisioning decisions required in the development of mission-critical voice applications for LTE.

SPEECH CODEC INTELLIGIBILITY TESTING IN SUPPORT OF MISSION-CRITICAL VOICE APPLICATIONS FOR LTE

Stephen D. Voran and Andrew A. Catellier[1]

We describe a major effort to quantify the speech intelligibility associated with a range of narrowband, wideband, and fullband digital audio coding algorithms in various acoustic noise environments. The work emphasizes the relationship between these intelligibility results and analogous ones for an analog FM land-mobile radio reference. The initial phase of this project includes 54 noise environments and 83 audio codec modes. We use an objective intelligibility estimator to narrow the scope and then design a practically sized modified rhyme test (MRT) covering 6 challenging yet relevant noise environments and 28 codec modes for a total of 168 conditions. The MRT used 36 subjects to produce 432 trials for each condition. Results show that intelligibility depends strongly on noise environment, data rate, and audio bandwidth. For each noise environment we identify codec modes that produce MRT intelligibility values that meet or exceed those of analog FM. We expect that these results can inform some of the design and provisioning decisions required in the development of mission-critical voice applications for LTE.

Keywords: ABC-MRT, acoustic noise, audio coding, background noise, MRT, speech coding, speech intelligibility

1. BACKGROUND

The National Public Safety Telecommunications Council (NPSTC) has defined seven high-level requirements for Mission-Critical Voice (MCV) networks for public safety [1]. Six of these requirements can be described as operating modes or system capabilities: Push-to-talk, Full Duplex, Group Call, Direct Mode, Talker Identification, and Emergency Alerting. The seventh requirement relates to audio quality and is actually a set of four quality-of-service thresholds:

> "Audio Quality: This is a vital ingredient for mission critical voice. The listener MUST be able to understand without repetition, and can identify the speaker, can detect stress in a speaker's voice, and be able to hear background sounds as well without interfering with the prime voice communications."

This report addresses this seventh requirement. The importance of audio quality is clear. As [1] indicates, if audio quality is sufficient, a listener can easily understand the messages without consistently asking for undue repetitions. The listener can additionally verify who is speaking, possibly gain understanding of the speaker's emotional state, and also understand the acoustic

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

environment in which the speaker is operating. All of these can help public safety practitioners work together to achieve critical goals in a timely way.

The same report [1] also prioritizes the four audio quality issues:

> "Audio Quality
> The transmitter and receiver audio quality must be such that, in order of importance:
> 1. The listener can understand what is being said without repetition.
> 2. The listener can identify the speaker (assuming familiarity with the speaker's voice).
> 3. The listener can detect stress in the speaker's voice, if present.
> 4. The background environment audio shall be sufficiently clear to the listener that sounds such as sirens and babies crying can be determined."

Indeed, the key issue here is speech intelligibility. Previous studies by the Public Safety Communications Research Program (PSCR), directly motivated by field reports from our public safety partners, have shown that speech intelligibility in background noise is the critical consideration [2]–[4].

Earlier PSCR work has investigated items two and three on the list: speaker identification and detection of speaker stress [5]–[7]. We tested speech intelligibility in parallel with the ability to identify a speaker from a set of speakers, and the ability to detect dramatized urgency in a speaker's voice. These tests were repeated across six communication systems with audio quality ranging from very good to very bad. As we moved from the best system to the worst, we found that speech intelligibility dropped off more rapidly than speaker identification performance or detection of urgency performance. In the context of these experiments at least, these results suggest that if a system preserves speech intelligibility, then it will also preserve the ability to identify speakers and detect urgency in speakers' voices. These results reaffirm that item one on the NPSTC list above is the critical issue.

Thus the PSCR has undertaken a detailed study of speech intelligibility for some of the various digital speech and audio codecs that can be used to provide voice over Long-Term Evolution (LTE) based radio networks. This study focuses on the speech intelligibility of transmissions produced by these codecs as a function of the acoustic noise environment. This report provides full details of the procedures and protocols used in the study and the results obtained.

The remainder of this section provides additional context for the study including discussion of the various factors that drive speech intelligibility. The following sections describe the audio codecs used in the study, production of digital speech and noise files for the study, objective estimation of speech intelligibility for these files, subjective testing of speech intelligibility using the Modified Rhyme Test (MRT), and analysis of the MRT results.

1.1 Speech Intelligibility Factors

We list four broad classes that organize the many factors that can affect speech intelligibility at the receive side:

- Acoustic noise at the transmit side and any noise mitigation techniques used there

- The audio codec used to encode speech and noise for transmission
- Impairments to the radio channel
- Acoustic noise at the receive side and any noise mitigation techniques used there

The factors of the first two classes are interrelated. They also appear first in the transmission chain. If these factors cause unintelligible speech to be transmitted, then the transmission will be unusable, regardless of how favorable the remaining factors are. Prior PSCR work regarding intelligibly of low-rate digital speech transmissions has demonstrated that the factors in these first two classes are indeed critical to successful communications [2]–[4]. Thus the present study addresses two key factors from these classes: acoustic noise at the transmit side and the choice of audio codec. Radio channel impairments can reduce speech intelligibility. Short-duration impairments can often be partially or completely concealed by robust coding and loss concealment algorithms. But impairments that persist for longer periods will often result in deterioration of intelligibility. The wide range of possible radio channel impairments and highly variable responses to those impairments place them outside the scope of the present study. How radio channel impairments might reduce the baseline intelligibility results from this study is an important topic for future study.

Noise mitigation techniques can improve speech quality in some cases. But improving speech intelligibility appears to be a bigger challenge [8]. Noise mitigation is extremely difficult in the severe noise cases that will drive the audio codec selection. Noise mitigation typically is not standardized, but rather is a factor that equipment manufacturers can use to gain proprietary advantage even while complying with encoding and transmission standards. In light of these observations, noise mitigation is outside the scope of the present study. How noise mitigation might improve the baseline intelligibility results from this study is a topic for potential subsequent study.

1.2 Speech Intelligibility Reference

The de facto reference point for mission-critical voice intelligibility is analog FM (AFM) transmission over land mobile radio (LMR). This reference point has been established through years of use and was thus used to judge the suitability of various Association of Public-Safety Communications Officials Project 25 (P25) offerings in various environments. While the latest digital P25 system offers many advantages compared to AFM over LMR (AFM-LMR), there are still important realistic high-noise environments where AFM-LMR has a significant speech intelligibility advantage over P25. This has justifiably caused some users to resist migrating away from AFM-LMR systems. But retaining such legacy systems is inefficient in many respects, including spectral use.

One solution is to seek LTE-based MCV applications that provide speech intelligibility that is no lower than that of AFM-LMR, even for the very difficult high-noise cases. By meeting this objective, such LTE applications can offer the public safety community the intelligibility that it is historically accustomed to in all cases and this will in turn address an important and warranted objection to migrating away from AFM-LMR systems. Note that the AFM-LMR speech

intelligibility reference is not a single value, but it is a set of values, one for each noise type and noise level of interest.

2. AUDIO CODECS

Audio codecs provide efficient (in terms of data rate) digital representations of audio signals. When the signal is speech alone a speech-specific signal model implemented in a speech codec can lead to efficient coding with good intelligibility. But when significant levels of background noise are combined with speech, broader or more robust signal models are required, and these may require higher data rates.

Note that "speech codec" indicates a codec optimized for speech alone, possibly with some level of robustness to background noises while "audio codec" often connotes a codec designed for arbitrary audio signals. The present study involves both speech and audio codecs. Since speech is a specific class of audio, we will often use the term "audio codec" instead of "speech or audio codec" for conciseness.

One key attribute of an audio codec is the audio bandwidth that it encodes and reproduces. The present study includes audio codecs with three different common nominal bandwidths. Narrowband (NB) audio codes support a nominal audio bandwidth that extends from approximately 300 Hz to 3.5 kHz. Wideband (WB) audio codecs support the range from approximately 50 Hz to 7 kHz. Fullband (FB) audio codecs have a nominal range from 20 Hz to 20 kHz. NB audio coding has been the historical standard for basic telecommunication services for many years. AFM-LMR is the reference system for this study, and it transmits using an audio bandwidth that is nominally NB. WB and FB represent enhancements that go beyond basic telecommunications and these are sometimes marketed as "High Definition Voice" or "HD-Voice." WB has been used in conferencing systems to improve the "realism" or "presence" of remote participants. WB is also emerging in commercial LTE voice services. In quiet conditions, uncoded WB speech is expected to have slightly higher intelligibility than uncoded NB speech [9]. FB is the bandwidth that one typically expects when listening to high-quality music recordings.

The P25 radio system requires audio coding at very low data rates (7.2 kb/s or less) and at the time of development this necessitated using NB speech. However, LTE can support much higher data rates. Thus LTE allows the opportunity to employ a much wider range of audio codecs. More specifically, LTE allows the opportunity to select WB or FB audio codecs and codecs that use coding paradigms that are less specific to speech and thus may be more robust to background noise.

Commercial LTE systems offer the Adaptive Multi-Rate (AMR) codec. This codec can operate at eight different data rates ranging from 4.75 to 12.2 kb/s. The WB AMR codec (AMR-WB) is emerging in LTE applications and it supports nine different data rates ranging from 6.6 to 23.85 kb/s. The 3GPP enhanced voice services (EVS) codec was standardized in September 2014. EVS offers NB, WB, and FB coding and offers rates that range from 5.9 to 128 kb/s. Other audio codecs can be used with LTE in what are typically called over-the-top applications.

The present study considers 83 different codec modes at the outset. The codecs are listed below, and all codec modes are enumerated in Table 1. The codecs and modes were selected to cover a range of data rates, speech bandwidths, and coding paradigms.

The P25 codec uses Advanced Multi-Band Excitation (AMBE). The software implementation used in this study was developed by and licensed from Digital Voice Systems Inc. of Westford, MA (dvsinc.com). The software is version 1.6 and represents the codec software that would be found in P25 digital radios that are currently on the market. This implementation was also used in our previous work [3], [4].

The G.711 Pulse Code Modulation (PCM) codec is specified in ITU-T Recommendation G.711. The software implementation used in this study is distributed as part of ITU-T Recommendation G.191 (itu.int).

The G.722 Adaptive Differential PCM (ADPCM) codec is specified in ITU-T Recommendation G.722. The software implementation used in this study is distributed as part of ITU-T Recommendation G.191 (itu.int).

The G.722.1 Modulated Lapped Transform (MLT) codec is specified in ITU-T Recommendation G.722.1 and the software implementation used in this study is distributed as part of that recommendation (itu.int).

The Adaptive Multi-Rate (AMR) codec is specified in the 3rd Generation Partnership Project (3GPP) TS 26.104. The software implementation used in this study is distributed as part of that technical specification (3gpp.org).

The Adaptive Multi-Rate Wideband (AMR-WB) codec is specified in 3GPP TS 26.204. The software implementation used in this study is distributed as part of that technical specification (3gpp.org).

The Enhanced Voice Services (EVS) codec was standardized by the 3GPP in September 2014. We used the latest available version for each step of this work. Thus the software implementation used to produce files for objective estimation of speech intelligibility (Section 4) is version 12.0.0. The software implementation used to produce files for the MRT (Sections 5 and 6) is version 12.2.0. Both implementations were provided as part of TS 26.442 (3gpp.org).

The Opus interactive speech and audio codec has been standardized by the IETF in RFC 6716 (ietf.org). The software implementation used in this study was built from libopus 1.1 (opus-codec.org).

The AAC-LD, AAC-ELD, and AAC-ELDsbr codecs have been standardized as parts of the ISO/IEC MPEG-4 standard (iso.org, mpeg.chiarglione.org). In this study these codecs were implemented through calls to the “afconvert” function that is part of Apple’s OS X 10.10 operating system.

Table 1. Audio codec modes considered in this study.

Codec Name	Audio Bandwidth	Target Data Rates (kb/s)	Number of Data Rates
P25 AMBE+2™	NB	2.45, 4.4	2
G.711 PCM	NB	64	1
G.722 ADPCM	WB	48, 64	2
G.722.1 MLT	WB	24, 32	2
AMR	NB	4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2, 12.2	8 (All available)
AMR-WB	WB	6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85	9 (All available)
EVS	NB	5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4	7
	WB	5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32, 48, 64	10
	FB	16.4, 24.4, 32, 48, 64	5
Opus	NB	5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32	8
	WB	5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32, 48, 64	10
	FB	16.4, 24.4, 32, 48, 64	5
AAC-LD	WB	32, 48	2
	FB	32,48,64	3
AAC-ELD	WB	32, 48	2
	FB	32,48,64	3
AAC-ELDsbr	WB	32	1
	FB	32, 48, 64	3
Total			83

In addition to the 83 codec modes, we also processed all signals through a software simulation of AFM transmission. This simulation was developed and verified through industry collaboration. It includes representative filters as would be used in typical analog transmitters and receivers, as well as pre-emphasis, deviation limiting, frequency modulation, demodulation, and de-emphasis. The simulation conforms to the specifications for bandwidth, deviation, frequency response, and sensitivity in the TIA-603 standard, and is described in significant detail in [3]. This simulation was used in our previous studies [3], [4].

Since radio channel impairments are outside the scope of the present study, the AFM simulation includes a perfect radio channel (12.5 kHz channel spacing). In other words, no interfering signals are simulated, and the simulation receiver is in the full-quieting state. Likewise, the audio codecs in this study are always operated with no bit errors and no packet losses between the encoder and the decoder.

3.SPEECH AND NOISE

3.1 Speech

The PSCR measures speech intelligibility using the Modified Rhyme Test (MRT). This selection was made in collaboration with public safety partners when the PSCR was formulating the project described in [2]. The selection stems from the protocol described in [10] for testing "face-to-face" voice communication through the mask associated with a self-contained breathing apparatus in noisy environments. This protocol specifies a testing environment and then specifies the MRT as the actual testing mechanism. The MRT is fully defined in [11]. In an MRT trial, a subject must identify the word presented from a set of six words that rhyme. At the conclusion of the MRT every condition under test receives an intelligibility rating that is based on the fraction of words correctly identified.

The MRT protocol specifies 50 sets containing 6 words each. 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 a 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 is presented in a carrier sentence: "Please select the word —." For example, when the test word is "bed," the carrier sentence is "Please select the word bed."

Two female and two male talkers were used to record the MRT words in the carrier sentence. Each is a native speaker of North American English. 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-isolated chamber with a noise criterion rating of NC-35 [12]. We used a studio grade microphone sampled at 48,000 smp/s, 16 b/smp for uncompressed direct-to-disk recording. Levels were set to eliminate clipping and low signal levels. These same recordings were used in previous PSCR studies [2]–[4]. Examples of the speech recordings can be heard and all source speech is available for download at PSCR.gov.

Other tests of speech intelligibility are available. Thus the intelligibility results presented in this report would be most precisely described as "MRT intelligibility" results. For conciseness we simply use the term "intelligibility" throughout this report.

3.2 Noise

We selected six types of acoustic background noise for consideration in this study. Table 2 provides descriptions of each type as well as a list of signal-to-noise ratios (SNRs) used with each. The noise types were selected to include a wide range of acoustic properties and to cover a range of public safety and civilian environments. The different noise types have different spectral and temporal properties and thus have different influences on speech intelligibility.

Each noise was recorded on location with professional-quality microphones and digital audio recorders that produced uncompressed recordings. Recordings used a sample rate of 48,000 smp/s and a minimum resolution of 16 b/smp. Levels were set to eliminate clipping and low signal levels. Some of these recordings were used in previous PSCR studies [2]–[4]. Examples of the noise recordings can be heard at PSCR.gov.

Note that there is no single "correct" or 'representative" SNR for any noise type. The SNRs measured in actual conditions will depend strongly on many environmental factors, including the physical relationships between the talkers, noise sources, and microphones. In addition, noise reduction techniques may influence the level and character of noises in a dynamic fashion. For each noise type we selected a range of SNR values. Our goal in this selection was to cover the range from unintelligible speech to fully intelligible speech.

Table 2. Noise environments considered in this study.

Name	**Description**	**SNRs (dB)**	**Number of SNRs**
Alarm	Alarm from firefighter's Personal Alert Safety System (PASS). Consists of a time-varying set of tones with noise power largely concentrated in the range 3150 to 3400 Hz.	-30, -25, -20 -15, -10, -5, 0, 5, 10, 15	10
Club	Sounds of crowd and live music recorded at nightclub.	-15, -10, -5, 0, 5, 10, 15	7
Coffee	Sounds of crowd, coffee preparation, and background music recorded at coffee shop.	-15, -10, -5, 0, 5, 10, 15, 20	8
Nozzle	Firefighting fog nozzle.	-20 -15, -10, -5, 0, 5, 10, 15, 20	9
Saw	K12 rescue saw cutting steel garage door.	-20 -15, -10, -5, 0, 5, 10, 15, 20, 25	10
Siren	Siren in yelp mode. Dominant power in 1 to 2 kHz range.	-30, -25, -20 -15, -10, -5, 0, 5, 10, 15	10
Total			**54**

3.3 Processing Speech and Noise

Starting with FB speech and noise recordings, we used the sample-rate conversion tools provided in [13] to convert each recording to the proper bandwidth and sample rate (8000 smp/s for NB or 16,000 smp/s for WB) for a given codec mode. We always selected the high-quality (brick-wall) low-pass filter option in these tools when converting to lower sample rates. Then for each version of each speech and noise recording (once for NB recordings, once for WB recordings and once for FB recordings), we computed a relative A-weighted level. We then calculated and applied an appropriate gain to the noise recording so it could be combined with the speech recording and achieve the desired SNR. The result was 54 noisy versions of each of the 1200 speech recordings for each audio bandwidth.

Finally we normalized the level of each recording to 28 dB below overload using the algorithm specified in [14]. These normalized recordings were then processed using the 83 codec modes

listed in Table 1 and using the AFM simulation as well. All processing was done in the digital signal processing domain. Examples of the resulting outputs can be heard at PSCR.gov.

4. OBJECTIVE ESTIMATION OF SPEECH INTELLIGIBILITY

As described in Section 3.1 above, the PSCR uses the MRT to evaluate speech intelligibility. This testing protocol requires significant investment of resources so we must carefully choose what material should be tested. More specifically, the 83 codec modes (see Table 1) and AFM simulation combined with the 54 noise environments (see Table 2) yield 4536 different conditions. Full MRT testing of this number of conditions is simply not feasible. (Section 5 describes a large scale MRT that covers 168 conditions.)

Thus this work includes a preselection stage where estimates of speech intelligibility are used to guide the application of MRT testing. These estimates come from the Articulation Band Correlation MRT (ABC-MRT) algorithm, previously developed and tested using PSCR MRT data [15]. The algorithm does not involve human listeners—it is a signal processing algorithm.

ABC-MRT uses MRT speech recordings and performs a speech recognition task that is analogous to the task in the MRT. But unlike most speech recognition algorithms, the ABC-MRT algorithm does not strive for maximal robustness to noise and distortion. Instead it strives for a robustness that is similar to that of human listeners. ABC-MRT uses temporal correlations within articulation index bands to select one of six possible words from a list. The rate of successful word identification becomes the measure of speech intelligibility, just as in the MRT. Because the robustness is similar to that of humans, the ABC-MRT scores are similar to those of humans. Thus we say ABC-MRT provides estimates of true MRT values.

For each trial (e.g., “Please select the word ‘bed’”) the ABC-MRT algorithm produces a single value, nominally in the range from zero to one. This value is analogous to those produced by the MRT. An output value of zero corresponds to no conveyance of speech information. Any values below zero are equivalent to zero and are indicative of estimation error. An output value of one corresponds to perfect conveyance of speech information. This value indicates that the correct word is identified on every trial. Any values above one are equivalent to one and are indicative of estimation error.

It is important to note that at present the ABC-MRT algorithm output is determined by the portion of the signal below 4 kHz. Thus, while ABC-MRT can process WB and FB signals, the algorithm output is mainly determined by the NB portion of the signal. While ABC-MRT showed high correlation (> 0.95) to MRT results across 139 NB conditions previously tested [15], many of the conditions in the present study have not been previously studied by ABC-MRT. Thus we must caution that past intelligibility estimation performance may not be an indicator of intelligibility estimation performance in this current, expanded environment. In short, we emphasize that ABC-MRT results are *estimates* of speech intelligibility, and we must be certain to treat them as *estimates*. Specifically, they do not stand alone but instead they guide the MRT work that will follow.

Figure 1 shows example ABC-MRT results for five NB codec modes and AFM in club noise. The means are calculated across 1200 trials (see Section 3.1). As expected, the intelligibility estimates consistently increase as SNR increases. Note that the two lowest rate codec modes (EVS at 5.9 kb/s and AMR at 4.75 kb/sec) generally show estimated speech intelligibility below

that of AFM. On the other hand the three higher rate codec modes generally show estimated speech intelligibility even with or above that of AFM.

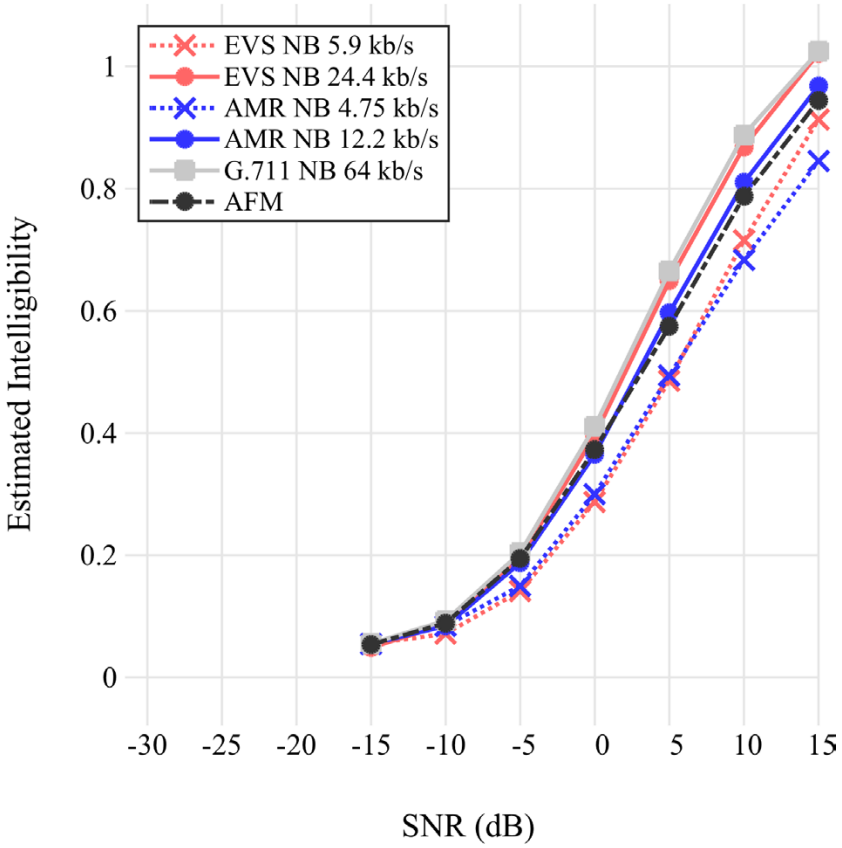

Figure 1. Estimated speech intelligibility example results for five NB codec modes and AFM in club noise.

Similarly, Figure 2 shows the same five codecs and AFM in siren noise. Here again, higher SNR's produce higher speech intelligibility estimates, and the effect of bit-rate is apparent as well.

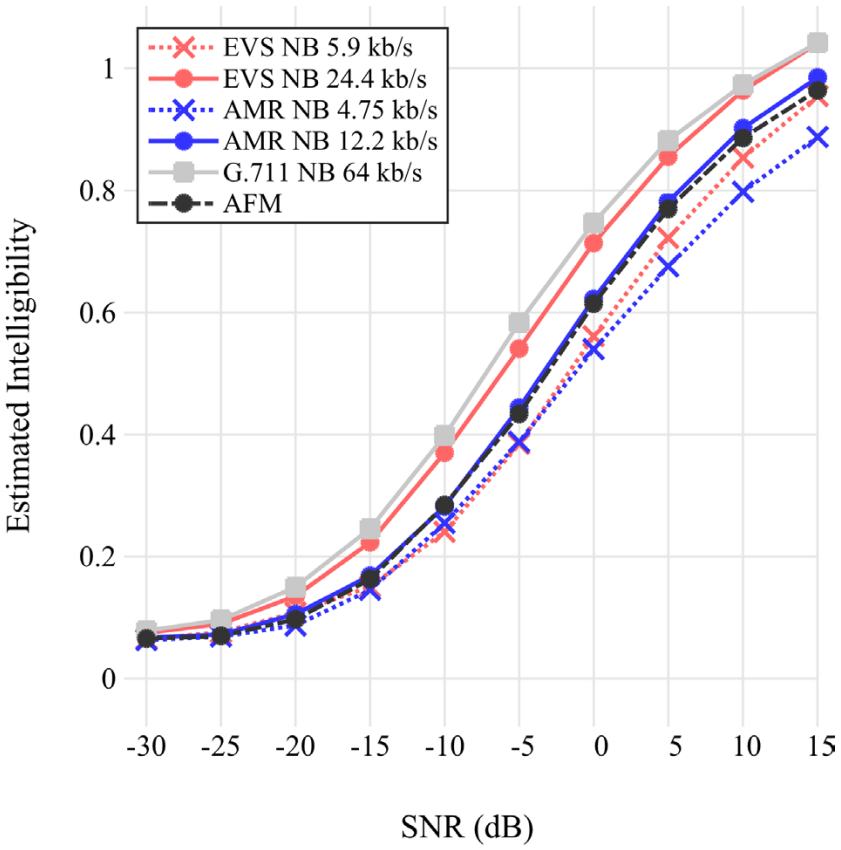

Figure 2. Estimated speech intelligibility example results for five NB codec modes and AFM in siren noise.

It is important to note again that these are example results for two noise types and that the plots show *estimated* speech intelligibility. It is not appropriate to draw any conclusions about the codec modes using either Figure 1 or Figure 2. Again, these results and others like them were produced to guide the design of the MRT.

4.1 Selecting SNRs

In order to make the problem space more tractable, the next step is to select a single SNR from each noise type to investigate more thoroughly. This SNR must be both demanding and relevant. Figure 1 and Figure 2 show that as SNR decreases, the noise dominates the intelligibility and the choice of codec mode eventually has very minimal influence on intelligibility. In the limit of very low SNR, noise completely obliterates the speech and intelligibility approaches zero, regardless of the codec mode selected. There is no motivation to allocate additional testing resources in these SNR regions because it is not possible for the codec modes to differentiate themselves.

As SNR is increased above these extreme values, codec mode can become a factor in speech intelligibility. To quantify the effect of codec mode in a rigorous way, we compare the mean estimated intelligibility of each codec mode with the mean estimated intelligibility of AFM using the use the t-test for the difference of means [16], [17]. More formally, let

$$\{\varphi_i(C)\}_{i=1}^{N} \tag{1}$$

and

$$\{\varphi_i(A)\}_{i=1}^{N} \tag{2}$$

be sets of ABC-MRT results for N=1200 trials made on codec mode C and AFM respectively. Then the sample means for C and AFM are given by

$$M_C = \frac{1}{N}\sum_{i=1}^{N}\varphi_i(C) \tag{3}$$

and

$$M_A = \frac{1}{N}\sum_{i=1}^{N}\varphi_i(A)\,, \tag{4}$$

respectively. The standard errors for M_C and M_A are given by

$$S_C = \frac{\sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(\varphi_i(C) - M_C)^2}}{\sqrt{N}} \tag{5}$$

and

$$S_A = \frac{\sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(\varphi_i(A) - M_A)^2}}{\sqrt{N}}, \tag{6}$$

respectively. Finally, the t-statistic is formed from a normalized difference of the means:

$$t = \frac{M_C - M_A}{\sqrt{S_C^2 + S_A^2}}. \tag{7}$$

When t <-1.96 one would typically conclude that M_C is lower than M_A with 95% confidence. The t-test for the difference of means is precise if the ABC-MRT results φ_i can be modeled as Gaussian random variables. When this is not the case, the test still quantifies the significance of the difference of two means, but the threshold t=-1.96 may not correspond to exactly 95% confidence. We note however that the t-test results are used only to select SNR's as described below. This selection process involves a minimization driven by a count of the numbers of cases with significant differences. In light of this application, we expect that the exact significance level for those differences (be it 95% or some slightly different value) will not influence the SNR selection process.

We use the t-test for the difference of means to classify the estimated intelligibility of each codec mode as either “lower than AFM” or “not lower than AFM.” Figure 3 shows the number of codec modes (out of 83 total) that are not lower than AFM as a function of SNR for siren noise. Note that the number of codec modes takes a minimum value of 61 when the SNR is 0 dB.

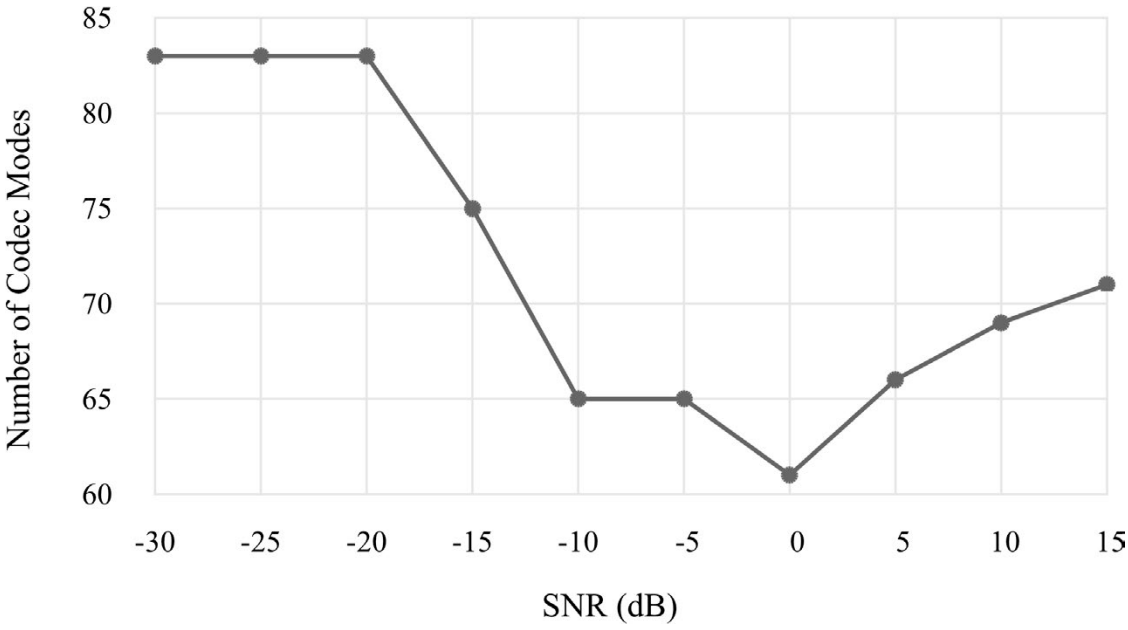

Figure 3. Number of codec modes that have estimated intelligibility not lower than AFM in siren noise.

We say that that 0 dB is the most demanding SNR for siren noise in this context because it minimizes the number of codec modes that meet our criterion (not lower than AFM) Below 0 dB, noise begins to dominate and the estimated speech intelligibility of the various codec modes begins converge to an equally low level. At -20 dB and below, none of the 83 codec modes receive an ABC-MRT score lower than AFM. As the SNR moves above 0 dB the lower noise level reduces the difficulty of the speech coding problem and an increasing number of codec modes can achieve the criterion of “not lower than AFM.”

Thus the SNR that produces a minimum in the number of codec modes that have estimated speech intelligibility not lower than AFM is a demanding and meaningful SNR. However, the selected SNR must also be relevant. For example, speech mixed with club noise with an SNR value of -5 dB produces the minimum number of codec modes with estimated speech intelligibility not lower than AFM. But Figure 1 indicates that at this SNR, AFM produces an estimated speech intelligibility that is less than 0.2. Further, across all 83 codec modes under consideration here, the highest estimated speech intelligibility found for this noise environment is only 0.215.

A simple and intuitive way to interpret an estimated speech intelligibility of 0.2 is that only 20% of the information is received correctly and that on average a message would have to be transmitted five times in order to communicate all of the information. This situation is of little

practical interest, and there is no motivation to apply further testing resources in this noise environment.

For purposes of SNR selection, we use an ABC-MRT score of 0.5 as the threshold for relevancy. If no codec mode achieves at least 0.5, then that SNR is not relevant. Thus we arrive at our full criteria for selecting a single SNR for each noise type. We select the SNR that minimizes the number of codec modes with estimated speech intelligibility not lower than AFM (the "demanding SNR" criterion), under the constraint that at least one codec mode must produce an estimated speech intelligibility of 0.5 or higher (the "relevant SNR" criterion).

Table 3 shows the SNR selected for each noise environment. The actual MRT results in Section 6 confirm that each of these is relevant—that is, actual MRT intelligibility results for AFM are above 0.5 for each of these noise environments. In addition, those results show that the relationship between SNR and intelligibility depends on the noise type. This is due to the diverse spectral and temporal characteristics represented by the noise types under consideration. The extremely low SNR selected for the Alarm noise does not result in a particularly low intelligibility for the AFM reference condition.

Input from the public safety community indicated that the Nozzle noise environment no longer holds significant interest (c.f. [2]). Thus we did not include Nozzle noise in the final MRT design. Instead we included the quiet environment (no noise added to the speech) which allows us to find the best-case intelligibility for each codec mode. Thus noise environments used in the MRT are Alarm, Club, Coffee, Saw, Siren, and Quiet.

Table 3. SNR selected for each noise type.

Name	**Description**	**SNR (dB)**
Alarm	Alarm from firefighter's Personal Alert Safety System (PASS). Consists of a time-varying set of tones with noise power largely concentrated in the range 3150 to 3400 Hz.	-30
Club	Sounds of crowd and live music recorded at nightclub.	+5
Coffee	Sounds of crowd, coffee preparation, and background music recorded at coffee shop.	+5
Nozzle	Firefighting fog nozzle. (Not used in MRT)	+5
Saw	K12 rescue saw cutting steel garage door.	0
Siren	Siren in yelp mode. Dominant power in 1 to 2 kHz range.	0
Quiet		

4.2 Selecting Codec Modes

Selecting a single SNR for each noise type reduces the number of conditions under consideration dramatically. But the practical constraints of MRT operations dictate that we also reduce the number of codec modes in the MRT. More specifically, given practical limitations of approximately 32 MRT subjects and about four hours of MRT time per subject, we find that around 168 conditions can be included in the MRT. This result follows from several

considerations: We require at least 350 trials per condition and this is driven by the statistical testing that will follow. We also know that on average, each trial requires about four seconds. We also have estimates for the time typically consumed by logistics, training, and breaks.

To allow for the most direct comparisons of codec modes, we elect to evaluate all selected codec modes in all six noise environments. This means that 28 codec modes can be selected ($28 \times 6 = 168$). From this point forward, we use the term "condition" to describe a combination of a codec mode and a noise environment. There are 168 conditions in this test.

As with the selection of noise environments, the codec mode selection is informed by the ABC-MRT results. For any codec type, ABC-MRT scores vary widely across the noise environments, as well as across any available audio bandwidths and data rates. As expected, ABC-MRT scores for a codec type generally increase as date rate is increased. But the data rate at which codec scores equate to AFM scores is a strong function of noise environment and is also influenced by audio bandwidth. These results guide us to the conclusion that the most useful MRT design cannot focus on a narrow range of data rates that are expected to produce an intelligibility near that of AFM. Instead, the data rates selected must cover most of the available range of data rates.

The final selection of codec modes balances the goal of including as many different codec types and bandwidths as possible against the goal of data rate inclusion described above, under the practical limitation that 28 codec modes can be selected. One of these "codec modes" is consumed by the need to include AFM (which is not a codec). In addition, to provide context for the most instructive data analysis, three of the "codec modes" must be the direct NB, WB, and FB conditions with no audio coding. In fact, only 24 of the 28 codec modes include an audio codec, so the term is used loosely here. The 28 codec modes chosen for the MRT are given in Table 4. The data rate for the uncoded modes is calculated from the native sample rate (8000, 16,000, and 48,000 smp/s) for NB, WB, or FB respectively, multiplied by the bit-depth of 16 b/smp.

Table 4. List of 28 codec modes with bandwidth and data rate.

Codec Mode Number	**Codec Type**	**Audio Bandwidth**	**Data Rate (kb/s)**
1	Analog FM	NB	NA
2	P25	NB	4.4
3	AMR	NB	5.9
4	AMR	NB	12.2
5	EVS	NB	5.9
6	EVS	NB	16.4
7	Opus	NB	5.9
8	Opus	NB	16.4
9	Uncoded	NB	128.0
10	AMR	WB	6.6
11	AMR	WB	15.85
12	AMR	WB	23.85
13	EVS	WB	5.9
14	EVS	WB	16.4
15	EVS	WB	32.0
16	Opus	WB	5.9
17	Opus	WB	16.4
18	Opus	WB	32.0
19	G.722.1	WB	24.0
20	G.722	WB	48.0
21	AAC-ELD	WB	32.0
22	Uncoded	WB	256.0
23	EVS	FB	16.4
24	EVS	FB	32.0
25	Opus	FB	16.4
26	Opus	FB	32.0
27	AAC-ELD	FB	32.0
28	Uncoded	FB	768.0

5. MODIFIED RHYME TESTING

5.1 Listening Lab

We conducted the MRT in two matched sound-isolated rooms with inside dimensions 305 cm long, 274 cm wide and 213 cm high (approximately 10 by 9 by 7 feet). In each room the floor is carpeted and all of the walls and the ceiling are covered with sound absorbing materials. Under normal conditions as would be experienced in the MRT, the noise level inside either room is below 26.5 dBA measured with a Brüel and Kjær Type 2250 sound level meter. When the air conditioning for a room is turned off, that level drops below 19.5 dBA for each room. These are extremely low noise levels and these measurements demonstrate that background noise is well-controlled in these labs.

Both rooms are configured so that the MRT subject sits on a chair in the center of the room behind a 76 cm by 152 cm (2.5 by 5 foot) work table. This table supports a loudspeaker, an LCD monitor screen, and a mouse as shown in Figure 4. Subjects were given the option to have the mouse and monitor positioned to the left of the speaker if preferred.

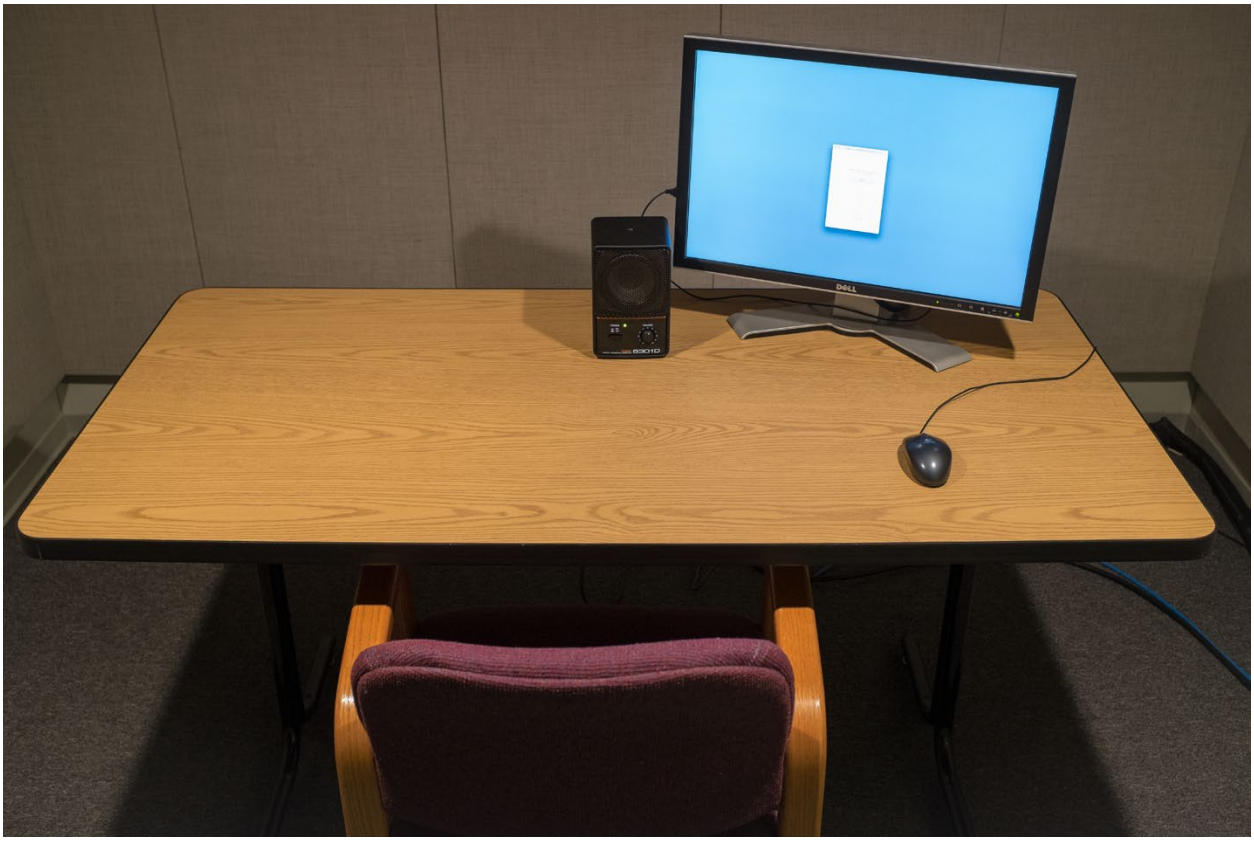

Figure 4. Photo depicting MRT lab setup.

As described in Section 3 the MRT recording format is digital files with 48,000 smp/s and 16 b/smp. The playback path includes a digital audio interface (USB to AES/EBU) so that the AES/EBU digital audio format is provided to the digital input of a Fostex Model 6301D Digital

Personal Monitor loudspeaker. Subjects were encouraged to adjust the volume knob on the front of this loudspeaker to achieve preferred listening level.

We used pink noise playback to characterize the combined frequency response of the playback electronics, the loudspeaker, and the room. Our spectral analysis was performed at the subject head location using octave-wide analysis bands (see ANSI S1.11). The composite response in the octaves centered at 125, 250, 500, 1000, 2000, 4000, 8000, and 16,000 Hz deviate no more than ±5 dB with respect to the response in the octave centered at 1000 Hz.

5.2 An MRT Trial

In the MRT a subject hears a carrier sentence (e.g., “Please select the word bed”) and then performs that task using a graphical user interface displayed on the LCD monitor screen. An image of the GUI presented on the screen is shown in Figure 5. The subject performs the task through a mouse click on the appropriate button. There are always six words to choose from and the order in which they appear (top to bottom) is randomized. This is an example of a forced-choice test from psychophysics. Once a button is clicked the next sentence is played, thus starting the next trial. It is not possible to replay any sentence.

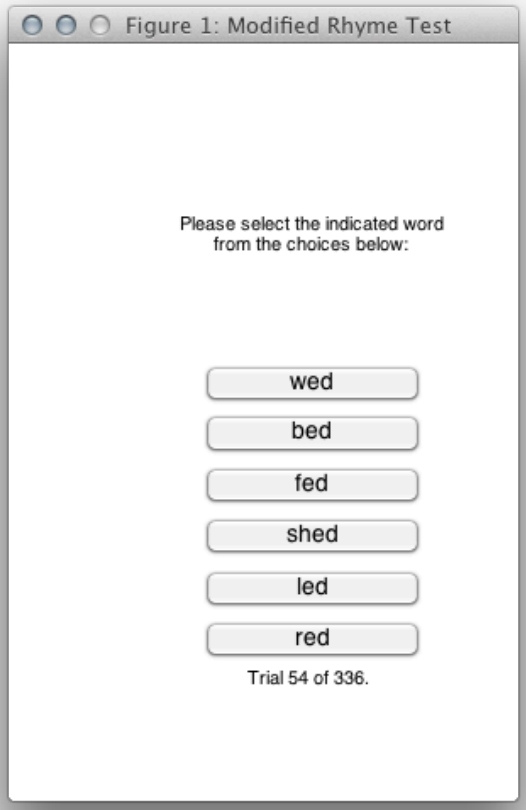

Figure 5. Screenshot of the MRT voting interface.

If the spoken sentence ends with the word “bed” and the recording has very high intelligibility, the word “bed” is easy to distinguish and the vast majority of the trials will lead to the selection of the correct answer. If the recording has very low intelligibility, subjects may hear “wed,” “fed,” “shed,” “led,” or “red” instead of “bed” and the vast majority of the trials will lead to the selection of an incorrect answer. MRT trials are performed repeatedly on each condition and each trial is classified as a success or a failure. This provides the raw data for further statistical analysis.

5.3 MRT Structure

The MRT consists of a practice session and six test sessions. The practice session contains ten trials that expose subjects to a range of noise environments, codec modes, and intelligibility levels. The practice session allows subjects to familiarize themselves with the MRT process and to resolve implementation issues before the actual test begins. In addition, the practice session allows the test administrator to confirm proper operation of the all equipment involved in the test

before any actual data are acquired. The data from the practice trials are not used in subsequent statistical analyses.

Each of six sessions is dedicated to a single noise environment. The relationship between session and noisy type is random and different for each of subjects 1 through 16. That relationship is inverted for subjects 17 through 32. For example subject 1 progressed through “coffee,” “siren,” “club,” “alarm,” “saw,” and “quiet.” Subject 17 heard these six sessions in the opposite order. The first session was “quiet” and the final session was “coffee.” This approach helps to balance the noise environment presentation order and thus minimize any effects in the MRT results related to noise environment presentation order.

A session length of 336 trials is consistent with the goal of keeping the total MRT time per subject near four hours. As described in Section 4.2, each session must cover all 28 codec modes, so this allows for 12 trials of each codec mode in every session ($12 \times 28 = 336$). These 12 trials are presented contiguously, so a single codec mode is heard 12 times before the test moves on to the next codec mode.

The 28 codec modes cover three different audio bandwidths: NB, WB, and FB. Changing bandwidths can induce additional perceptual effects that could conflate with the speech intelligibility results of interest here. This motivates us to minimize the magnitude and number of the bandwidth changes. The magnitude of the bandwidth changes is minimized by allowing changes between NB and WB as well as WB and FB, while prohibiting changes between NB and FB. The number of bandwidth changes is held to a minimum value of two changes per session by grouping all codec modes for each of the three bandwidths. The result of these two policies is that in any MRT session the trials cover all NB codec modes, then all WB codec modes, and finally all FB codec modes, or the reverse order (all FB, then all WB, then all NB). In other words, every session uses an increasing or decreasing bandwidth progression.

For 8 of the 32 subjects, 8 bandwidth progressions (increasing or decreasing) are randomly chosen and within each bandwidth a random order of the applicable codec modes is selected. Using these 8 orderings, another 8 orderings are created by reversing the codec mode order within each bandwidth. These 16 orderings are then doubled to 32 orderings by reversing the bandwidth progression of each. This approach helps to balance the bandwidth presentation order and the codec mode presentation order and thus minimize any effects in the MRT results that might stem from these two properties.

The MRT produces 384 trials (32 subjects × 12 trials) for each condition. Conditions can be most directly compared if all other variables are held fixed. This uniformity could be achieved by repeating the same 12 test sentences for every subject and every condition. But this would create an extremely repetitive MRT environment (each sentence would be heard 168 times) and this can lead to excessive subject fatigue and erroneously low performance. It would also test each codec mode with a vanishingly small sampling of speech signals (just 12 sentences out of the 1200 available). This leads to results that are extremely sensitive to the choice of those twelve sentences rather than robust and representative results that we seek.

Our MRT design allows much less repetition and allows each codec mode to be tested with 384 different sentences. To achieve this design we first select a fixed set of 384 sentences. For

maximal breadth of test material we use MRT recordings from two different female talkers and two different male talkers. The MRT specification includes 50 lists of 6 words. To allocate these we consider the first half (lists 1–25) and the second half (lists 26–50). We allocate MRT lists 1 through 16 to female 1 (the first 16 lists of the first half). We allocate MRT lists 26 through 41 to female 2 (the first 16 lists of the second half). Similarly we allocate MRT lists 10 through 25 to male 1 (the last 16 lists of the first half) and MRT lists 35–50 to male 2 (the last 16 lists of the second half).

Since each list contains 6 words, 16 lists allow the use of 96 MRT sentences. We create a different random order of these 96 sentences for each talker and for each of the 168 conditions. For each condition, subject 1 hears the first 3 sentences of each list. Since there are 4 talkers this is a total of 12 sentences (12 trials) per condition, as desired. Subject 2 hears sentences 4, 5, and 6 from each list, and subject 32 hears sentences 94, 95, and 96. The sentences from each talker are heard consecutively, but a different random talker order is selected in every case. For example, on one condition subject 1 may hear 3 sentences from Male 1, then 3 sentences from Female 2, then 3 from Female 1, and finally 3 sentences from Male 2. On the next condition subject 1 might hear 3 sentences from Female 2, then 3 from Female 1, then 3 from Male 1, and finally 3 sentences from Male 2.

By this design, when 32 subjects have completed the MRT, each condition has been tested with the same 384 sentences (96 from each of 4 talkers). In addition, each subject has heard 12 randomly selected sentences for each of 168 conditions. Any subjects beyond the initial 32 simply repeat the work of the first 32. Subjects 1 and 33 hear exactly the same material in exactly the same order. The same is true for subjects 2 and 34, 3 and 35, and 4 and 36.

The randomization processes used in this MRT design prevent repeatable patterns in the presentation of trials, and thus we can be certain that the subjects' answers were based solely on their ability to understand the key word in each trial. The balancing processes used in this MRT design help to make aggregate results across subjects more immune to presentation order effects or presentation position effects.

5.4 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 using radio links in noisy environments thus making their MRT results especially relevant to the question at hand.

We recruited the participation of 36 subjects from the public safety community. Each subject traveled to our laboratory facility in Boulder, Colorado in June, July, or August of 2015. Subjects were from various locations spanning the U.S. Each subject reported his or her professional experiences. Dispatch or other communication focused activities were reported 17 times, fire service experience was reported 16 times, law enforcement experience was reported 9 times, EMS or paramedic activities were reported 6 times, and one subject reported disaster management experience. This is a total of 49 professional experience areas, resulting in an average of 1.4 experience areas per subject

Subjects also reported total years of service. The mean value is 19.3 years and the median value is 20.5 years. Twenty-seven of the subjects (75%) are male and nine are female. For the 27 males the estimated ages are distributed as follows: 2 aged 20–29 years, 6 aged 30–39, 11 aged 40–49, 6 aged 50–59, and 2 aged 60 and up. The estimated age distribution for the 9 females is: 3 aged 30–39, 4 aged 40–49, and 2 aged 50–59. The median of the estimated age bin is the 40s for both males and females.

According to standard protocol for experiments with human subjects, each subject read and signed a statement of informed consent. Next each subject read a set of written MRT instructions. Key points from these instructions include an invitation to "adjust the volume to your preferred listing level as often as you wish," an invitation to "position the speaker, mouse and monitor for best use," and two requests to "turn off your phone for every session 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 began with the practice session, as described in Section 5.3. 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. The administrator offered a break after each session. Subjects typically elected to take a five or ten minute break after every second or third session.

The majority of the test sessions lasted between 20 and 28 minutes. This corresponds to an average of 3.6 to 5.0 seconds per MRT trial. The mean session length is 22.7 minutes, corresponding to 4.1 seconds per trial. When considering time used in introductory procedures, training, and breaks, the typical total time per subject was indeed near four hours.

6. ANALYSIS AND DISCUSSION

6.1 Number and Distribution of Trials

We designed the MRT to achieve exact balances in several respects when 32 subjects participate. To allow for the inevitable cancellations we recruited more than 32 subjects. While some cancellations did occur, we ultimately were fortunate to have the participation of 36 subjects.

The 36 subjects each completed 2016 trials (after the practice session) for a grand total of 72,576 trials. The MRT achieved exact balance of talker gender: 36,288 trials with female talkers and the same number with male talkers.

Each condition received 12 trials from each of 36 subjects for a total of 432 trials. The first 384 of these trials (produced by the first 32 subjects) are associated with the exact same 384 recordings for each of the 168 conditions. These 384 recordings include all 300 words (6 words from each of 50 lists) at least once. Fourteen lists are used twice, and this accounts for an additional 84 words. The 432 trials are perfectly balanced across the four talkers (108 trials from each talker).

In this MRT design the six noise environments were assigned to the six sessions differently for each subject using a balanced approach. The median (calculated across all 36 subjects) position (1 to 6) of every noise environment is 3, 3.5, or 4. This indicates very good balance between early and late positioning for every noise environment.

Similarly, the 28 codec modes were assigned positions in the sessions using a balanced approach. The median (calculated across all 6 noise environments and 36 subjects) position (1 to 336) of each codec mode within a session is 168.5 in every case. This indicates exact balance between early and late positioning for every codec mode.

6.2 MRT Data Analysis

The MRT is simply a 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 [18]. 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 parameter p.

For any group of N trials resulting in S 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 [18]:

$$\hat{p} = \frac{S}{N}. \tag{8}$$

That is, the estimated probability of success in the underlying Bernoulli model is simply the fraction of successes observed. For every condition tested we have $N = 432$ trials. Table 5 gives the number of successful trials for each condition tested.

Table 5. Number of successful trials (out of 432 total trials) for each condition.

Codec Mode Number	Description	Saw	Club	Coffee	Siren	Alarm	Quiet
1	Analog FM NB	264	317	326	361	362	417
2	P25 NB 4.4	206	253	274	337	311	398
3	AMR NB 5.9	234	291	295	331	184	410
4	AMR NB 12.2	274	314	336	353	231	419
5	EVS NB 5.9	252	285	298	321	230	412
6	EVS NB 16.4	285	337	337	363	321	421
7	Opus NB 5.9	251	282	292	310	309	397
8	Opus NB 16.4	268	335	335	364	345	419
9	Uncoded NB	288	341	340	373	367	422
10	AMR WB 6.6	287	333	329	371	265	413
11	AMR WB 15.85	303	354	363	391	333	421
12	AMR WB 23.85	311	373	372	391	353	421
13	EVS WB 5.9	257	315	323	351	249	421
14	EVS WB 16.4	289	370	367	388	318	418
15	EVS WB 32	316	376	356	400	337	423
16	Opus WB 5.9	278	318	322	354	284	397
17	Opus WB 16.4	314	361	366	386	319	424
18	Opus WB 32	315	381	375	394	339	428
19	G.722.1 WB 24	321	378	367	396	352	422
20	G.722 WB 48	311	379	377	405	329	424
21	AAC-ELD WB 32	311	365	382	389	370	420
22	Uncoded WB	326	374	373	401	370	425
23	EVS FB 16.4	329	371	359	401	303	421
24	EVS FB 32	321	381	385	394	334	426
25	Opus FB 16.4	290	366	365	389	303	428
26	Opus FB 32	311	372	368	401	342	421
27	AAC-ELD FB 32	290	344	342	385	370	418
28	Uncoded FB	331	392	373	418	380	426

This estimated probability of success provides the basis for reporting intelligibility. Because the MRT offers six word choices, the expected lower limit for the probability of success is one-sixth (0.167). In other words, even with the speech signal turned off (clearly a case of zero intelligibility), any subject could select the correct word one-sixth of the time on average simply

by selecting one of the six word options at random. Thus [11] specifies a transformation that maps $\hat{p}$ to intelligibility, denoted by R:

$$R = \frac{6}{5}\left(\hat{p} - \frac{1}{6}\right). \tag{9}$$

This relationship maps $\hat{p} = \frac{1}{6}$ (the success rate for guessing) to $R = 0$. It also maps $\hat{p} = 1$ to $R = 1$, as desired. Note that the uncertainty in the estimate $\hat{p}$ and thus in R is properly accounted for in Section 6.5. Table 6 gives value of R for each condition tested.

Table 6. Intelligibility (R) for each condition ($0 \leq R \leq 1$).

Codec Mode Number	Description	Saw	Club	Coffee	Siren	Alarm	Quiet
1	Analog FM NB	0.533	0.681	0.706	0.803	0.806	0.958
2	P25 NB 4.4	0.372	0.503	0.561	0.736	0.6634	0.906
3	AMR NB 5.9	0.450	0.608	0.619	0.719	0.311	0.939
4	AMR NB 12.2	0.561	0.672	0.733	0.781	0.4412	0.969
5	EVS NB 5.9	0.500	0.5912	0.628	0.692	0.439	0.944
6	EVS NB 16.4	0.592	0.736	0.736	0.808	0.6912	0.969
7	Opus NB 5.9	0.497	0.583	0.611	0.661	0.658	0.903
8	Opus NB 16.4	0.544	0.7301	0.731	0.811	0.758	0.964
9	Uncoded NB	0.600	0.747	0.744	0.836	0.819	0.972
10	AMR WB 6.6	0.597	0.725	0.714	0.831	0.536	0.947
11	AMR WB 15.85	0.642	0.783	0.808	0.886	0.725	0.969
12	AMR WB 23.85	0.664	0.836	0.833	0.886	0.7801	0.969
13	EVS WB 5.9	0.514	0.675	0.697	0.775	0.492	0.969
14	EVS WB 16.4	0.603	0.828	0.819	0.878	0.683	0.961
15	EVS WB 32	0.678	0.844	0.789	0.911	0.736	0.975
16	Opus WB 5.9	0.572	0.683	0.694	0.783	0.5889	0.903
17	Opus WB 16.4	0.672	0.8023	0.817	0.872	0.686	0.978
18	Opus WB 32	0.675	0.858	0.842	0.894	0.742	0.989
19	G.722.1 WB 24	0.692	0.850	0.819	0.900	0.778	0.972
20	G.722 WB 48	0.664	0.8523	0.847	0.925	0.714	0.978
21	AAC-ELD WB 32	0.664	0.814	0.861	0.881	0.828	0.967
22	Uncoded WB	0.706	0.8389	0.836	0.914	0.828	0.981
23	EVS FB 16.4	0.714	0.831	0.797	0.914	0.642	0.969
24	EVS FB 32	0.692	0.858	0.869	0.894	0.728	0.983
25	Opus FB 16.4	0.606	0.817	0.814	0.881	0.642	0.989
26	Opus FB 32	0.664	0.833	0.822	0.914	0.750	0.969
27	AAC-ELD FB 32	0.606	0.756	0.750	0.869	0.828	0.961
28	Uncoded FB	0.719	0.889	0.836	0.961	0.856	0.983

6.3 Analog FM Reference

Figure 6 shows the *R* values for the six conditions that include AFM, organized from lowest intelligibility to highest intelligibility. The saw noise environment produces the lowest *R* value (near 0.53), the club and coffee shop noises provide similar and somewhat higher *R* (in the neighborhood of 0.70). The alarm and siren noises also provide similar *R* results and these are in the neighborhood of 0.80. As expected, the quiet condition produces the highest intelligibility (*R* is near 0.96).

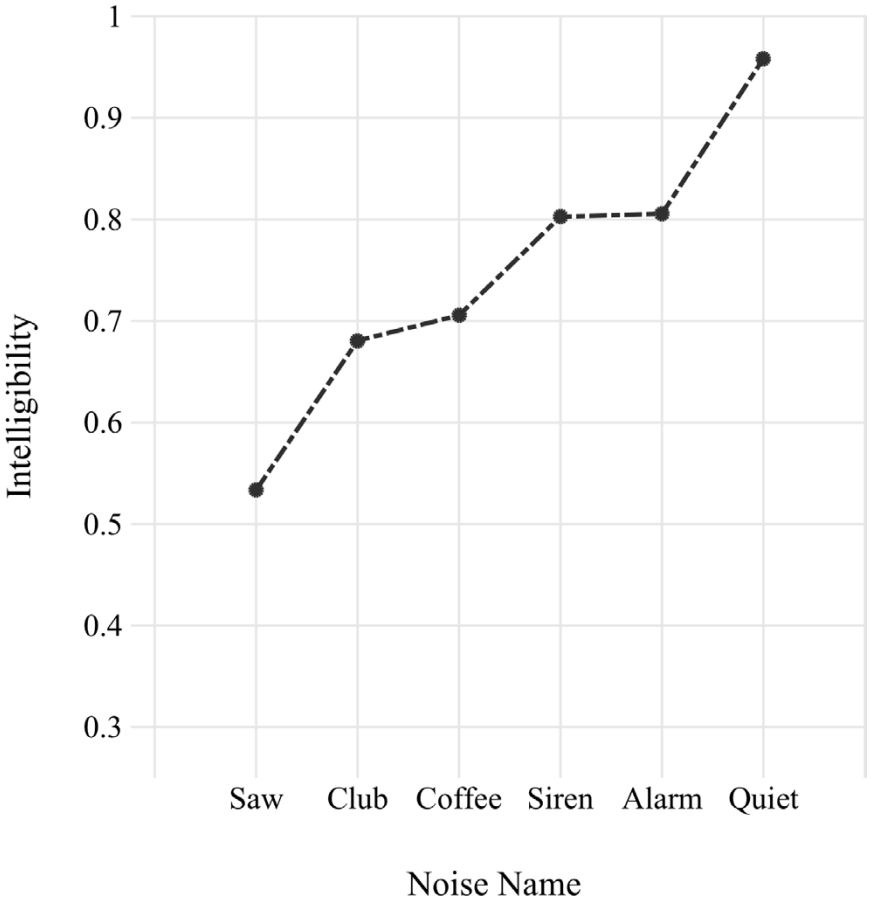

Figure 6. AFM intelligibility for each noise environment.

These results cover a usable range of intelligibility levels. The lowest of these is produced by the saw noise environment. Here *R* is near 0.5 and on average each message will have to be transmitted twice in order to successfully convey the required information. While this may be annoying, it is certainly not a futile endeavor. Since these mixtures of speech and noise all produce usable results through AFM, the corresponding results through the other codec modes are indeed interesting and relevant.

6.4 Other Codec Modes

Figures 7–12 show the intelligibility (*R*) for the 28 codec modes in the 6 noise environments. The figures include a dashed black line that shows the *R* value for the AFM reference in the specified noise environment and a dashed green line that shows the *R* value for the uncoded condition. For visual clarity we show mean values but not confidence intervals. The uncertainty in the MRT results is properly accounted for in Section 6.5. These figures show NB, WB, and FB codec modes separately and for each audio bandwidth *R* is plotted as a function of codec data rate. These figures show that the various noise environments produce significant variation, but the general trends are as expected: increasing bit rate and increasing bandwidth generally lead to higher intelligibility.

For each audio bandwidth the intelligibility results for the uncoded condition represent upper limits for intelligibility at that bandwidth. In quiet, moving from NB to WB produces a modest increase in intelligibility and there is only an insignificant increase when moving from WB to FB. Different noise environments cause substantial variation in the magnitudes of these two bandwidth-driven intelligibility increases.

In some cases codec modes have produced *R* values greater than the *R* value for the uncoded condition with the corresponding bandwidth. But comparison testing that accounts for the uncertainty in these MRT results (analogous to the comparisons described in Section 6.5) shows that none of these differences are statistically significant.

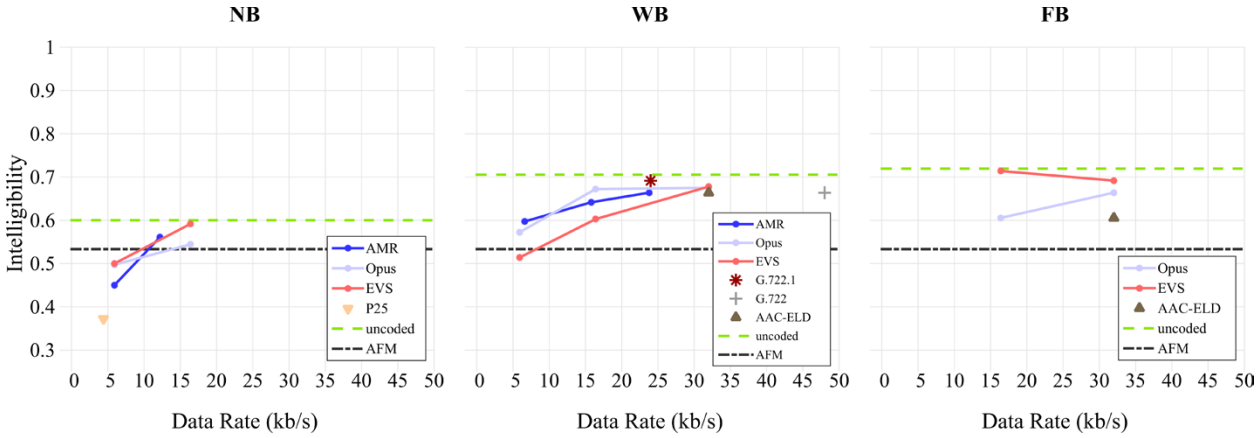

Figure 7. Intelligibility vs. data rate for all 28 codec modes in saw noise environment.

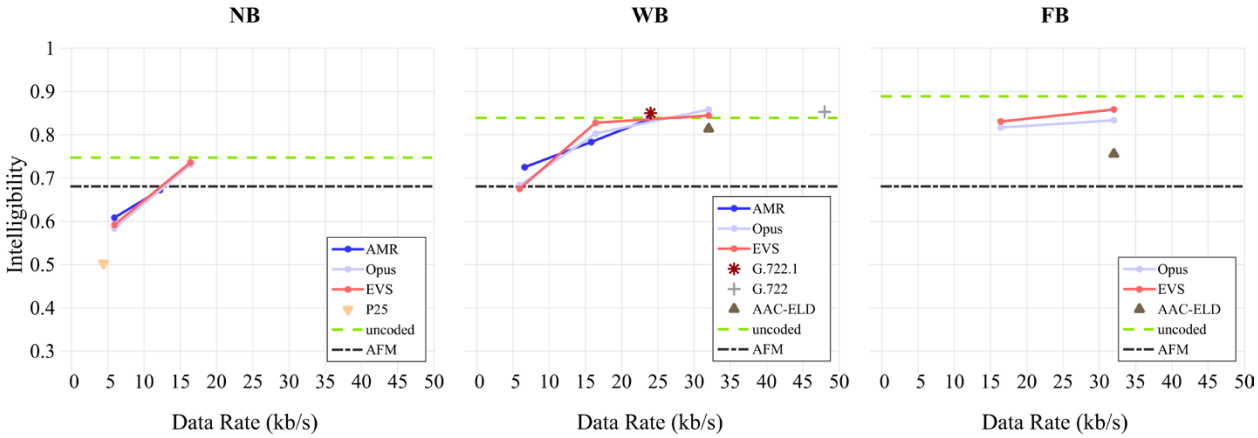

Figure 8. Intelligibility vs. data rate for all 28 codec modes in club noise environment.

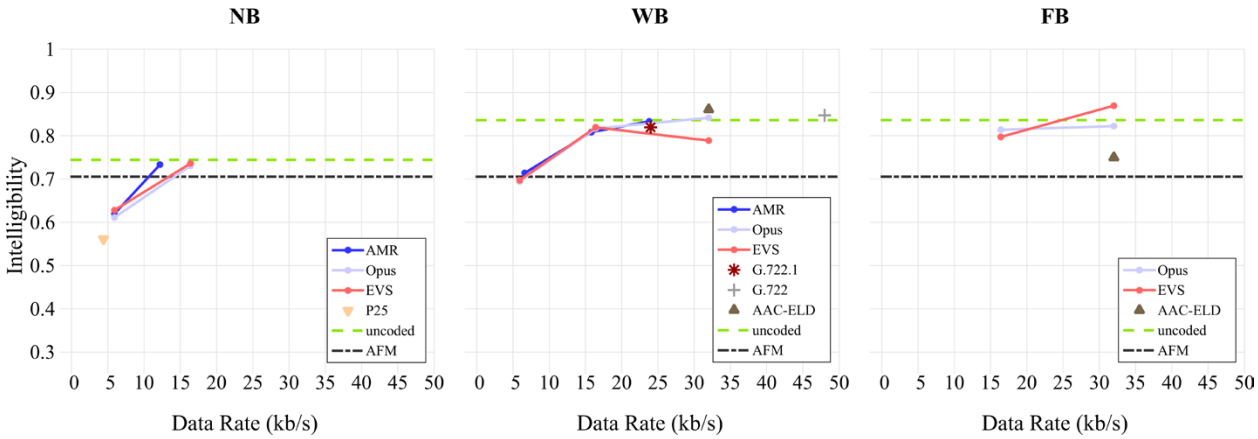

Figure 9. Intelligibility vs. data rate for all 28 codec modes in coffee noise environment.

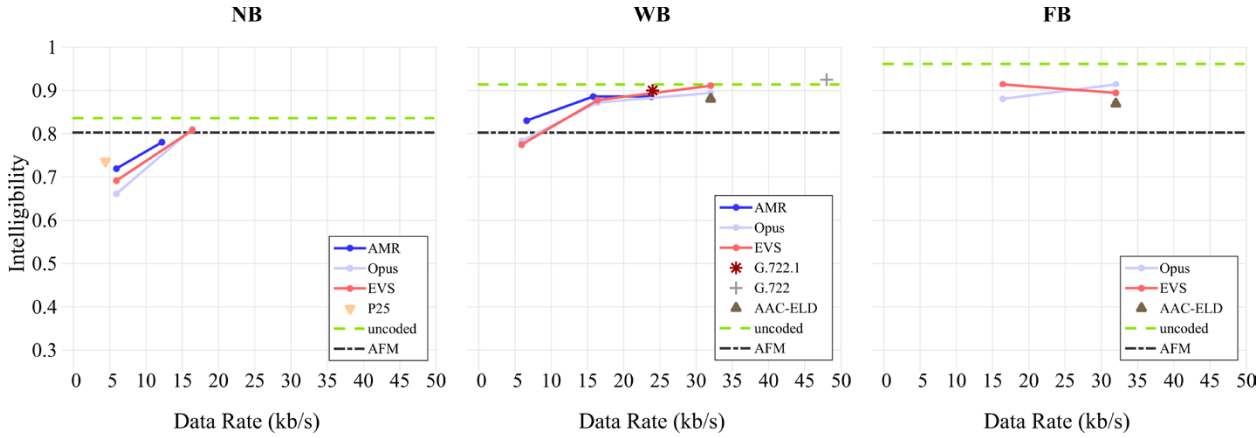

Figure 10. Intelligibility vs. data rate for all 28 codec modes in siren noise environment..

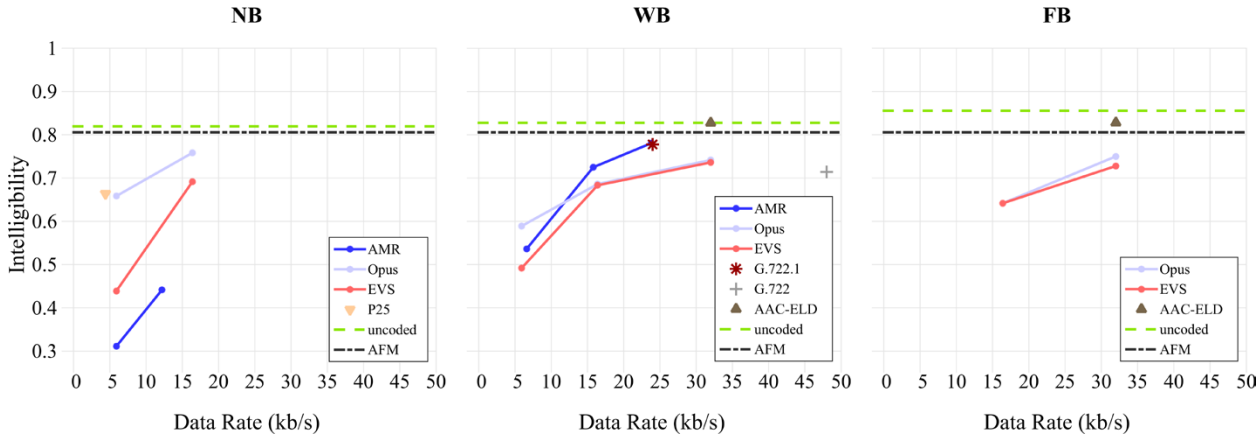

Figure 11. Intelligibility vs. data rate for all 28 codec modes in alarm noise environment.

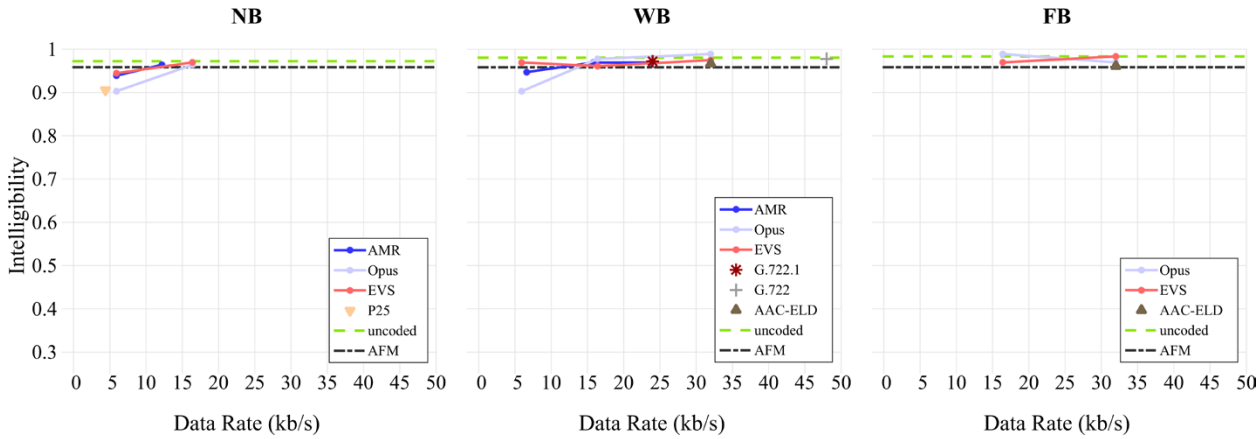

Figure 12. Intelligibility vs. data rate for all 28 codec modes in quiet environment.

6.5 Comparisons

Figures 7–12 show intelligibility of codec modes and allow for easy visual comparison with AFM. Each intelligibility value shown is based on a large but finite number of Bernoulli trials and thus has some inherent uncertainty. As with all work of this type, we must carefully consider the question of statistical significance. Thus we will posit a null hypothesis and perform statistical tests to determine when we should reject the null hypothesis.

For any noise environment we can tabulate successes and failures for any given codec mode (denoted by *C*) alongside of those for AFM (denoted by *A*) as shown in Table 7.

Table 7. Example table comparing Codec Mode C with AFM.

	Number of Successes	**Number of Failures**	**Total**
Codec Mode C	S_C	F_C	N_C
AFM	S_A	F_A	N_A
Total	$S_C + S_A$	$F_C + F_A$	$N_C + N_A$

We can apply the chi-squared test for independence of categorical data [16], [17], [19] to test for independence of these numbers of successes with respect to the row variable (codec mode *C* vs. AFM). Thus the null hypothesis is "the success rates shown in the two rows are independent of the labeling of the rows." In other words, codec mode *C* and AFM do not have statistically significantly different success rates.

To apply the chi-squared test for independence of categorical data we form the chi-squared statistic from the normalized squared deviations between the observed results and the expected results under the null hypothesis. Note that since this MRT is balanced, $N_C = N_A = N$ and this simplifies the expressions that follow. The expected results are easily extracted from the totals given in the table:

$$S_{NULL} = \frac{S_C + S_A}{N_C + N_A} N = \frac{S_C + S_A}{2}\,, F_{NULL} = N - S_{NULL}. \tag{10}$$

Next we form the chi-squared (χ^2) statistic associated with the two-by-two core of Table 7:

$$\chi^2 = \frac{(S_C - S_{NULL})^2}{S_{NULL}} + \frac{(S_A - S_{NULL})^2}{S_{NULL}} + \frac{(F_C - F_{NULL})^2}{F_{NULL}} + \frac{(F_A - F_{NULL})^2}{F_{NULL}}$$

$$= 2\left(\frac{(S_C - S_{NULL})^2}{S_{NULL}} + \frac{(F_C - F_{NULL})^2}{F_{NULL}}\right). \tag{11}$$

This χ^2 statistic has one degree-of-freedom ((number of rows − 1) × (number of columns − 1)). Table 8 shows the value of the χ^2 for every condition in the MRT compared to AFM. As expected, the statistic takes the value zero when AFM is compared to itself.

Table 8. Values of the chi-squared (χ^2) statistic for testing the null hypothesis.

Codec Mode Number	Description	Saw	Club	Coffee	Siren	Alarm	Quiet
1	Analog FM NB	0	0	0	0	0	0
2	P25 NB 4.4	15.6955	21.1179	14.7491	4.2951	17.4826	7.8103
3	AMR NB 5.9	4.2662	3.7525	5.5022	6.5331	157.6645	1.3836
4	AMR NB 12.2	0.4926	0.0529	0.6461	0.5163	92.2640	0.1476
5	EVS NB 5.9	0.6929	5.6094	4.5231	11.1372	93.4913	0.7444
6	EVS NB 16.4	2.2033	2.5164	0.7845	0.0341	11.7485	0.6345
7	Opus NB 5.9	0.8124	6.6677	6.5697	17.3530	18.7407	8.4914
8	Opus NB 16.4	0.0783	2.0252	0.5216	0.0772	2.2495	0.1476
9	Uncoded NB	2.8896	3.6715	1.2842	1.3039	0.2195	1.0298
10	AMR WB 6.6	2.6502	1.5901	0.0568	0.8942	54.7068	0.4899
11	AMR WB 15.85	7.8038	9.1335	9.8098	9.2325	6.1864	0.6345
12	AMR WB 23.85	11.4853	22.5679	15.7785	9.2325	0.6569	0.6345
13	EVS WB 5.9	0.2369	0.0236	0.0557	0.7983	71.3689	0.6345
14	EVS WB 16.4	3.1398	19.9588	12.2561	7.3124	13.3688	0.0357
15	EVS WB 32	14.1832	25.3798	6.2647	16.7657	4.6820	1.5429
16	Opus WB 5.9	0.9703	0.0059	0.0988	0.3974	37.3262	8.4914
17	Opus WB 16.4	13.0665	13.2641	11.6145	6.1786	12.8189	2.1887
18	Opus WB 32	13.6185	30.5429	18.1552	11.4332	4.0000	6.5116
19	G.722.1 WB 24	17.1990	27.3717	12.2561	13.0668	0.8067	1.0298
20	G.722 WB 48	11.4853	28.4039	19.8551	22.2825	7.8708	2.1887
21	AAC-ELD WB 32	11.4853	16.0376	24.5319	7.9225	0.5723	0.3441
22	Uncoded WB	20.5445	23.4822	16.5481	17.7860	0.5723	2.9851
23	EVS FB 16.4	22.7152	20.8066	7.6736	17.7860	22.7270	0.6345
24	EVS FB 32	17.1990	30.5429	27.6476	11.4332	5.7931	3.9532
25	Opus FB 16.4	3.4009	16.7806	10.9931	7.9225	22.7270	6.5116
26	Opus FB 32	11.4853	21.6761	12.9183	17.7860	3.0682	0.6345
27	AAC-ELD FB 32	3.4009	4.6940	1.6894	5.6535	0.5723	0.0357
28	Uncoded FB	24.2323	44.2240	16.5481	42.3943	3.0924	3.9532

The chi-squared statistic measures the deviation of the outcomes for AFM and Codec Mode *C* from the outcome that is expected under the null hypothesis. It goes to zero as outcomes for AFM and Codec Mode *C* converge and it gets larger as they diverge. The cumulative distribution function of this statistic is well-characterized [16], [17], [19] and it is thus known that when the null-hypothesis is true, the statistic will exceed 3.841 less than 5% of the time.

In much of science and engineering, it is common to reject the null hypothesis when the probability of rejecting it erroneously is less than 5%. This is sometimes described as a 95% significance or confidence level. We follow this practice. When $3.841 < \chi 2$ we reject the null hypothesis. Table 9 uses the equal sign with no shading to indicate conditions where the null

hypothesis has not been rejected. The plus sign with light yellow shading indicates that the null hypothesis has been rejected and the condition has an R value that exceeds that of AFM. The minus sign with light blue shading indicates that the null hypothesis has been rejected and the condition has an R value lower than that of AFM. (Note that the deterministic and monotonic relationship between $\hat{p}$ and R given in (9) allows us to map the outcomes of $\hat{p}$-domain hypothesis tests to the R domain.) The final row and column of Table 9 tabulate the total number of cases where intelligibility is lower than AFM by codec mode and by noise environment.

Four of the "codec modes" were included in the MRT for reference purposes and are not truly codecs. These are AFM and the three uncoded conditions. As expected, none of the uncoded conditions fail to produce intelligibility equivalent to or better than AFM. Crossing the remaining 24 codec modes with the 6 noise environments results in 144 non-reference conditions. In 34 of these 144 conditions we have found that the intelligibility is lower than that of AFM.

The final row of Table 9 shows that over half of these cases are associated with the alarm noise environment. This environment does not present an extraordinary challenge to AFM (see Figure 6) but it does present significant challenges for many of the codec modes considered here (see Figure 11). As noted earlier, the alarm noise contains an attention-grabbing time-varying set of tones positioned near the upper edge of the NB passband. Much of the alarm noise power is between 3150 and 3400 Hz and that region contains a very small portion of the speech power. Investigation shows that AFM tends to attenuate the signal in this region, thus reducing the alarm component without much effect on the speech. More generally, intelligibly in the alarm noise environment may be fairly sensitive to frequency response at the upper edge of the NB passband.

It follows that simple fixed filtering at codec inputs might boost intelligibility in alarm noise. While the MRT results for the alarm noise environment are correct, it may be that they deserve lesser weight in light of the potential for fairly simple mitigation. This situation is in stark contrast to other noise environments where noise power is spread across much of the speech spectrum and a simple attenuation of the higher frequencies will not have the incidental effect of improved intelligibility.

Table 9. Hypothesis test outcomes for 168 conditions. A minus sign with light blue shading indicates intelligibility lower than AFM, an equal sign with no shading indicates intelligibility the same as AFM, and a plus sign with light yellow shading indicates intelligibility higher than AFM.

Codec Mode Number	Description	Saw	Club	Coffee	Siren	Alarm	Quiet	Number of –
1	Analog FM NB	=	=	=	=	=	=	0
2	P25 NB 4.4	-	-	-	-	-	-	6
3	AMR NB 5.9	-	=	-	-	-	=	4
4	AMR NB 12.2	=	=	=	=	-	=	1
5	EVS NB 5.9	=	-	-	-	-	=	4
6	EVS NB 16.4	=	=	=	=	-	=	1
7	Opus NB 5.9	=	-	-	-	-	-	5
8	Opus NB 16.4	=	=	=	=	=	=	0
9	Uncoded NB	=	=	=	=	=	=	0
10	AMR WB 6.6	=	=	=	=	-	=	1
11	AMR WB 15.85	+	+	+	+	-	=	1
12	AMR WB 23.85	+	+	+	+	=	=	0
13	EVS WB 5.9	=	=	=	=	-	=	1
14	EVS WB 16.4	=	+	+	+	-	=	1
15	EVS WB 32	+	+	+	+	-	=	1
16	Opus WB 5.9	=	=	=	=	-	-	2
17	Opus WB 16.4	+	+	+	+	-	=	1
18	Opus WB 32	+	+	+	+	-	+	1
19	G.722.1 WB 24	+	+	+	+	=	=	0
20	G.722 WB 48	+	+	+	+	-	=	1
21	AAC-ELD WB 32	+	+	+	+	=	=	0
22	Uncoded WB	+	+	+	+	=	=	0
23	EVS FB 16.4	+	+	+	+	-	=	1
24	EVS FB 32	+	+	+	+	-	+	1
25	Opus FB 16.4	=	+	+	+	-	+	1
26	Opus FB 32	+	+	+	+	=	=	0
27	AAC-ELD FB 32	=	+	=	+	=	=	0
28	Uncoded FB	+	+	+	+	=	+	0
Number of –		2	3	4	4	18	3	

The results shown in Table 9 are presented differently in Figure 13 (for the 144 non-reference conditions only). This presentation supports visualization of the data rate and audio bandwidth parameters. Data rate increases monotonically (but not uniformly) as we move from left to right in this display. Audio bandwidth increases from NB to WB and then FB as we move up. The colors of Table 9 are used again in this display and can be used to judge the success of any codec

mode in terms of meeting or exceeding the intelligibility of AFM. As success is achieved in more and more noise environments we see fewer light blue squares and more white or even light yellow squares. This presentation allows us to easily see, at the level of individual noise environments, the pros and cons associated with changing codec types, data rates, audio bandwidths, or combinations of these factors.

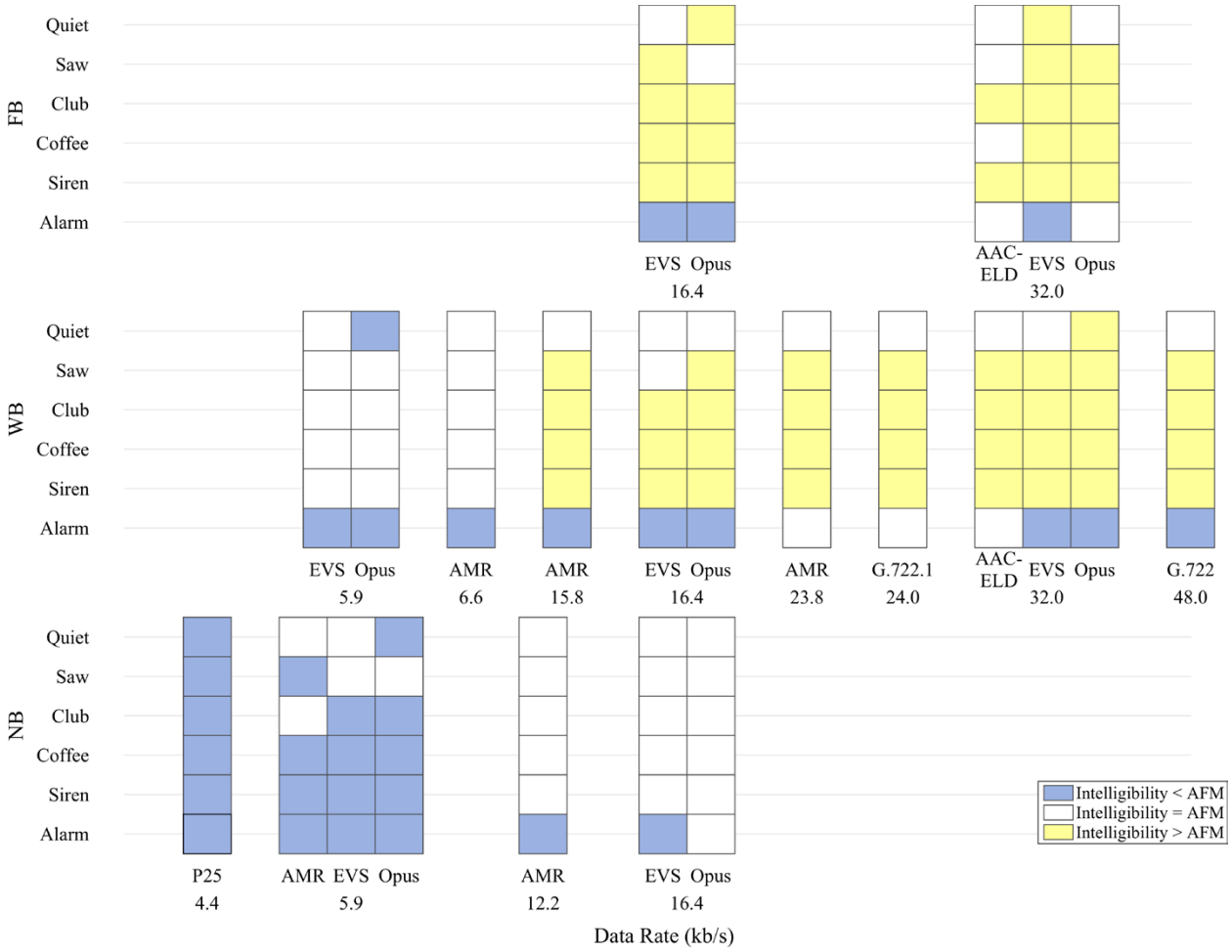

Figure 13. Hypothesis test outcomes for 24 non-reference codec modes organized by increasing data rate and audio bandwidth. Light blue indicates intelligibility lower than AFM. White indicates intelligibility the same as AFM. Light yellow indicates intelligibility higher than AFM.

7. CONCLUSIONS

The PSCR has completed a study of the speech intelligibility for some of the digital speech and audio codecs that could potentially be used to provide mission-critical voice communications over LTE-based radio networks. The study focuses on speech intelligibility in some of the harsh noise environments that may be experienced by public safety practitioners. We have provided detailed descriptions of the design, implementation, analysis, and results of the study.

First we offer several conclusions related to the work process itself. We adopted a two-phased approach and this was an innovation and a success. More specifically, the initial phase of this work considered 83 codec modes and 54 noise environments (4482 total conditions). We successfully applied an objective estimator of speech intelligibility to evaluate these conditions and thereby reduced the set considered in the second phase to 28 codec modes and 6 noise environments (168 conditions). This reduction in size allowed us to design a practically sized MRT. We then conducted the MRT and found that subjects could complete 2016 trials in about 4 hours. With a total of 36 subjects, this is a total of 72,576 trials and these were evenly distributed at 432 trials per condition.

Conclusions related to speech intelligibility follow directly from analysis of these MRT trials. Specifically, the MRT intelligibility results are shown in Figures 7–12. Table 9 and Figure 13 give the results of the statistical tests that compare each codec mode with the AFM reference in every noise environment. These figures and tables allow one to draw a huge number of very specific conclusions as a function of codec type, data rate, bandwidth, and noise environment.

Finally, we provide some broader conclusions drawn from those very specific statistical tests. In the quiet environment we observe that digital speech coding is very effective from an intelligibility perspective. Table 9 shows that only three codec modes at the very lowest rates (4.4 and 5.9 kb/s) fail to match the intelligibility of AFM.

Table 9 also shows that only six codec modes produce intelligibility no lower than AFM in all six of the noise environments. The data rates for these six range from 16.4 to 32 kb/s. These codec modes include one NB mode, three WB modes, and two FB modes. We can also ask which codec modes produce intelligibility no lower than AFM in at least five of the six noise environments. Here the result jumps from six to nineteen. More specifically three of the seven NB code modes, eleven of the twelve WB codec modes, and all five of the five FB codec modes meet this standard. The corresponding data rates range from 6.6 to 48 kbps.

Our work also makes it apparent that when higher bit rates are available the use of WB and FB coding can provide superior intelligibility in many noise environments. This intelligibility can exceed that of AFM intelligibility in many environments. In fact, Table 9 and Figure 13 show that every WB or FB codec mode that uses 15.85 kb/s or greater delivers intelligibility higher than that of AFM in either two, three, four, or five of the six noise environments.

Overall, we conclude that there are multiple audio coding options that can deliver speech intelligibility that meets or exceeds that of the typical analog FM system used in public safety communications, even in the context of multiple diverse and harsh public safety noise environments. And as expected, the success of various audio coding options clearly depends on the data rate available for transmission of the digitally coded audio.

8. REFERENCES

[1] NPSTC, "Mission critical voice communications requirements for public safety," Littleton, CO, 2011.

[2] D. Atkinson and A. Catellier, "Intelligibiliity of selected radio systems in the presence of fireground noise: Test plan and results," NTIA Report 08-453, Washington D.C., 2008.

[3] D. Atkinson and A. Catellier, "Intelligibility of analog FM and updated P25 radio systems in the presence of fireground noise: Test plan and results," NTIA Report 13-495, Washington D.C., 2013.

[4] D. Atkinson, S. Voran and A. Catellier, "Intelligibility of the adaptive multi-rate speech coder in emergency-response environments," NTIA Report 13-493, Washington D.C., 2013.

[5] S. Voran, "Listener detection of talker stress in low-rate coded speech," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, 2008.

[6] A. Catellier and S. Voran, "Speaker identification in low-rate coded speech," in *Proc. 7th International Measurement of Audio and Video Quality in Networks Conference*, Prague, 2008.

[7] A. Catellier and S. Voran, "Relationships between intelligibility, speaker identification, and the detection of dramatized urgency," NTIA Report 09-459, Washington D.C., 2008.

[8] P. Loizou and G. Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 47-56, 2011.

[9] S. Voran, "Listener ratings of speech passbands," in *Proc. 1997 IEEE Workshop on Speech Coding for Telecommunications*, Pocono Manor, PA, 1997.

[10] NFPA, "NFPA 1981 Standard on open-circuit self-contained breathing apparatus (SCBA) for emergency services," Quincy, MA, 2007.

[11] ANSI, "ANSI S3.2 American national standard method for measuring the intelligibility of speech over communication systems," New York, 1989.

[12] L. Beranek, Cirteria for noise and vibration in communities, buildings, and vehicles," in *Noise and vibration control*, New York, McGraw-Hill, 1971, pp. 564-566.

[13] ITU-T, "Rec. P.191: Software tools for speech and audio coding standardization," Geneva, 2012.

[14] ITU-T, "Rec. P.56: Objective measurement of active speech level," Geneva, 2011.

[15] S. Voran, "Using articulation index band correlations to objectively estimate speech intelligibility consistent with the modified rhyme test," in *Proc. 2013 IEEE International*

Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2013.

[16] E. L. Crow, F. A. Davis and M. W. Maxfield, Statistics Manual, New York: Dover, 1960.

[17] A. M. Mood, F. A. Graybill and D. C. Boes, Introduction to the Theory of Statistics, New York: McGraw-Hill, 1974.

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

[19] R. V. Hogg and E. A. Tanis, Probability and Statistical Inference, New York: Macmillian, 1977.

ACKNOWLEDGEMENTS

Funding for this work was provided by the DHS Science and Technology Directorate, Cuong Luu, Program Manager. The work was conducted by the PSCR, Andrew Thiessen and Dereck Orr, Program Managers. This work builds on recordings, protocols, and lab capabilities created by the late DJ Atkinson as part of the earlier PSCR studies. The present work would not have been possible without DJ's vision, leadership, and hard work. We are deeply indebted to DJ, and we seek to honor his memory through this present work.

Kathy Mayeda performed all of the travel planning, authorization, logistics, and reimbursement functions that allowed 36 members of the public safety community to participate in the MRT. Her contributions were critical to the success of this project and we offer her our sincere thanks. We also thank Dylan Hicks for administering MRTs and his support in MRT administration.

We also extend our thanks to the many members of the public safety community who traveled to the PSCR Laboratories in Boulder, Colorado, to participate in the MRT. This report would not be possible without this very generous support from those individuals and their supporting agencies. Finally, we recognize the technical reviewers who provided essential input to this report and we extend deep gratitude to ITS Publications Officer Lilli Segre for her tireless and thorough editorial revisions that have produced this final product.

NTIA FORM 29
(4-80)

U.S. DEPARTMENT OF COMMERCE
NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO.	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE		5. Publication Date
		6. Performing Organization Code
7. AUTHOR(S)		9. Project/Task/Work Unit No.
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
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.)

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

abstract; appendix; conclusion; document; figures; format; heading; introduction; policy; references; style guide; tables

17. AVAILABILITY STATEMENT	18. Security Class. (This report)	20. Number of pages
☒ UNLIMITED.	Unclassified	61
☐ 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.

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.