NISTIR 8206

Mission Critical Voice QoE Mouth-to-Ear Latency Measurement Methods

Jesse Frey Jaden Pieper Tim Thompson

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8206



NISTIR 8206

Mission Critical Voice QoE Mouth-to-Ear Latency Measurement Methods

Jesse Frey Jaden Pieper Tim Thompson Communications Technology Laboratory

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8206

February 2018



U.S. Department of Commerce Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Acknowledgments

The authors would like to acknowledge the technical guidance and enthusiasm for testing audio-related performance in radio systems provided by Steve Voran of the Institute for Telecommunication Sciences. The authors would also like to acknowledge the references and assistance provided by Amanda Koepke for the proper reporting of the uncertainty of the collected data. The authors would finally like to acknowledge the technical contributions in making proper field measurements provided by Don Bradshaw of the National Institute of Standards and Technology's (NIST) Public Safety Communications Research (PSCR) group.

Abstract

Mouth-to-ear (M2E) latency describes the time it takes speech input in a voice communication transmit device to be output from a receiving device, and has been identified as a key component of quality of experience (QoE) in communications. NIST's PSCR division developed a method to measure and quantify the M2E latency of any communications system transmitting audio, with specific emphasis on push to talk (PTT) devices. This measurement method is the first step in establishing QoE key performance indicators (KPI) for mission critical voice (MCV) and a measurement system to quantify these QoE KPIs. Additional measurement methods will be established and published in the near future.

The measurement system provides a fair platform for the comparisons of M2E latency across radio communications technologies. Both single and two location measurement systems were developed. The single location measurement system is a simpler setup ideal for measurements performed in a single, controlled setting. The two location system allows for the measurement of M2E latency between devices in two distinct locations and adds the capability to see potential effects of distance and signal propagation on the latency a user experiences. Example measurements of the M2E latency of VHF and UHF land mobile radios (LMR) operating in both direct mode and in trunked mode were performed. These tests demonstrated that both the single and two location tests return consistent measurement results.

Key words

Audio; Communications; Delay; Direct Mode; Handset; Key Performance Indicator (KPI); Land Mobile Radio (LMR); Latency; Mission Critical Push To Talk (MCPTT); Mission Critical Voice (MCV); Mouth-to-ear (M2E); Project 25 (P25); Public Safety; Push To Talk (PTT); Quality of Experience (QoE); Quality of Service (QoS); Repeater; Trunked Mode; Ultra High Frequency (UHF); Very High Frequency (VHF).

Table of Contents

1	Introduction		
2	Res	earch Summary	1
3	Background		1
	3.1	Mouth-to-Ear Latency	2
	3.2	LMR Radio Systems	3
4	Technical Approach		3
	4.1	General Test Information and Setup	4
	4.2	Delay Calculations and Audio Samples	7
	4.3	Maximizing Speech Quality	8
		4.3.1 Ground Loop Effects	8
		4.3.2 Audio Volume	9
	4.4	Microcontroller	10
	4.5	Measurement Characterizations	11
	4.6	Uncertainty Calculations	13
5	Exa	mple Delay Measurements	15
	5.1	Additional Measurement Information	15
	5.2	Measurement Characterization	17
	5.3	UHF Direct Mode	17
	5.4	UHF Trunked Mode	19
	5.5	VHF Direct Mode	19
	5.6	VHF Trunked Mode	22
6	Con	clusion	22
Re	References		

List of Tables

Table 1 Summary of example M2E latency measurement results	2
Table 2 Test thinning information	17

List of Figures

Fig. 1	MCPTT access time and M2E latency	3
Fig. 2	LMR radio system diagrams	4
Fig. 3	Diagram of single location test	5
Fig. 4	Two location test diagrams	6
Fig. 5	Example audio envelope and windows	8
Fig. 6	Circuit schematic for the optoisolator	11
Fig. 7	Microcontroller for controlling the PTT button on the radio	12

Fig. 8 Diagram of single location system characterization	12
Fig. 9 Diagram of two location measurement characterization	13
Fig. 10 Two location field measurement sites	16
Fig. 11 Adapted LMR headsets	16
Fig. 12 Measurement characterization results	18
Fig. 13 Raw UHF direct mode measurements	20
Fig. 14 Raw UHF trunked mode measurements	21
Fig. 15 Raw VHF direct mode measurements	23
Fig. 16 Raw VHF trunked mode measurements	24

1. Introduction

Public safety communications devices experience delays which affect the Quality of Experience (QoE) for first responders relying on them to perform their jobs. These delays must be quantified to determine the performance of current land mobile radio (LMR) technology and how new capabilities, particularly broadband technologies, compare. Mouth-to-ear (M2E) latency is a fundamental delay component in all communication systems, including LMR and Long Term Evolution (LTE) systems, and is defined as the time between speech input into one device and its output through another.

Mouth to ear latency is an accepted and widely used term and we adopt it for this report. However we will point out "microphone-to-loudspeaker" would actually be more descriptive of these measurements. True mouth to ear latency would have to include the acoustic path from the mouth to the microphone, the electronic path from the microphone to the loudspeaker, and then the acoustic path from the loudspeaker (or earpiece) to the ear. The initial and final acoustic paths are unknown and variable, they will increase latency by roughly one ms per foot, and they will likely not exceed several feet (or ms) in most cases.

Here a device agnostic M2E latency measurement system is introduced and some example measurements for some LMR radio systems are provided. The intention of the measurement system is to provide a platform on which fair comparisons of M2E latency can be made across communication technologies and devices.

2. Research Summary

Two M2E latency measurement systems were developed for this project. One is a single location system that requires both the transmit and receive audio devices to be in close proximity, and the other is a two location system that allows for the devices to be separated. Each system has an associated latency that must be measured so that device measurements can be corrected. These systems work to provide a fair platform for the comparisons of M2E latency across radio communications technologies.

Example measurements were performed using a variety of LMR system configurations. M2E latency was measured for both very high frequency (VHF) and ultra high frequency (UHF) radios in direct and trunked modes using a single and two location measurement system in a lab setting. Two location measurements were also performed over the air in a field test. The single and two location measurement systems were shown to provide consistent results. These results and their associated uncertainties can all be seen in Table 1.

3. Background

NIST's Public Safety Communications Research (PSCR) group held a roundtable event in March 2017 with industry and public safety representatives to identify expectations and metrics that would enable PSCR to understand, measure, monitor, and predict Mis-

	Single Location Lab	Two Location Lab	Two Location Field
	[ms]	[ms]	[ms]
Measurement	21.85 ± 0.07	21.85 ± 0.07	21.85 ± 0.07
Characterization			
UHF-Direct	201.4 ± 0.4	201.2 ± 0.3	201.8 ± 0.4
UHF-Trunked	415.8 ± 2.8	413.1±3.3	417.0 ± 2.9
VHF-Direct	201.7 ± 0.5	201.6 ± 0.4	202.4 ± 0.4
VHF-Trunked	403.9 ± 1.8	403.3 ± 2.8	405.3 ± 1.2

Table 1. Summary of example M2E latency measurement results. Device measurements corrected for measurement system delay.

sion Critical Voice (MCV) QoE across LMR, LTE, and future technologies. Quantifying QoE is a departure from traditional Quality of Service (QoS) metrics which focus on network and device performance. QoE focuses on the end users and their experience with the communications system. This project will quantify and expand existing voice quality measurement capabilities to be used as a generally-applicable QoE key performance indicator (KPI)/measurement. In particular, the measurement system presented here is based around an audio in/audio out approach, in order to create a device independent system that allows for direct measurement comparisons across a variety of communications technologies. This project is the first of a series of measurement methods focused on establishing and quantifying MCV QoE-based KPIs. Future projects will further develop QoE KPIs and their associated measurement systems.

Researchers are encouraged to replicate the M2E latency measurement system developed by PSCR engineers. Data gathered during the measurement experiments are available at https://doi.org/10.18434/T4/1422492 and the MATLAB code developed for the measurement system is available at https://github.com/usnistgov/mouth2ear. Limited technical assistance is available by contacting Tim Thompson at (303) 497-6613 or at tim. thompson@nist.gov.

3.1 Mouth-to-Ear Latency

Some of the factors affecting QoE are the delays experienced in push-to-talk (PTT) call setup and reception. In LMR trunked radio systems and future Mission Critical PTT (MCPTT) systems, the time required for the system to grant a channel from the PTT request is referred to as PTT Access Time. PTT Access Time is KPI 1 in Fig. 1. The delay experienced from the talker's audio capture and when the audio is received and played back is referred to as the M2E latency. M2E latency is KPI 3 in Fig. 1.

The 3rd Generation Partnership Project (3GPP) defines M2E latency (KPI 3) as "the time between an utterance by the transmitting user, and the playback of the utterance at the

¹3GPP Technical Specification (TS) 22.179 MCPTT over LTE Stage 1 (Release 14)



Fig. 1. Illustration of MCPTT access time and M2E latency¹

receiving user's speaker" [1]. The testing covered in this report focuses on M2E latency. A future testing effort will focus on quantifying PTT Access Time (KPI 1).

The Project 25 Telecommunications Industry Association (TIA) standard for Digital C4FM/CQPSK Transceiver Measurement Methods defines transmitter throughput delay and receiver throughput delay for individual transmitters and receivers within the LMR system [2]. The M2E latency for direct mode testing would simply be the combination of transmitter and receiver throughput delays. For a trunked mode system the transmitter and receiver throughput delays of the repeater would also contribute to the overall M2E latency.

3.2 LMR Radio Systems

The M2E latency measurement methods presented here were initially designed for PTT technology, but are viable for any audio communications system. The example measurements using the system focused on available PTT LMR systems. M2E latency was measured and quantified for VHF and UHF radio systems deployed at PSCR in Boulder, CO in both direct mode and trunked mode. Direct mode does not use any radio infrastructure and transmits directly from one handset to another. Trunked mode operation involves a LMR repeater which processes the request to transmit by monitoring a control channel, then assigns a specific frequency for the LMR to use to communicate with the other LMR handset(s). These modes are depicted in Fig. 2.

4. Technical Approach

A useful M2E latency measurement system must be device independent to allow for direct comparison across a variety of communication devices and services. Thus the system presented here is an audio in/audio out approach, which directly compares transmitted audio to received audio. This system is universal for all communication systems transmitting audio, and requires minimal effort to interface with. In particular, devices must be outfitted with a cable that can receive the transmit audio from the audio interface, and output received audio to an input on the audio interface. The required connections for these cables are specific to the audio interface used. An example of this is shown in Fig. 11. Both a single and



Fig. 2. LMR radio system diagrams

two location measurement system are presented here. The two location system primarily was designed for future research and it makes it possible to demonstrate that propagation is generally a negligible contributor to M2E latency.

In order to achieve consistent results when comparing radio technologies, measurements should be used on systems communicating via cabled RF. This ensures the minimum delay specification is reported without real channel conditions causing additional delays or increasing variability.

4.1 General Test Information and Setup

The measurement system involves using a computer running MATLAB to both play and record audio out of an audio interface. The system simultaneously sends audio to a transmit device while recording either received audio or a timecode output from a timecode generator. A microcontroller is attached to the computer via USB and to the transmit device to control PTT functionality.

In particular, the single location test setup requires one laptop, a microcontroller, and one audio interface, as well as the two communication devices on which the M2E latency measurement is being performed. The computer sends audio out via USB to the audio interface which then outputs the sound, through a playback jack, to the transmit device. The microcontroller is used to toggle when the device transmits. The receive device outputs the received audio to the audio interface where it is then recorded in MATLAB. The recording and audio playback start simultaneously. Thus the delay measurement consists of both the delay of the communication devices and the inherent time it takes for the signal to travel through the measurement system. The characterization of the measurement system delay is discussed in a later section, and allows for measurements to be corrected to only reflect the delay of the communication devices. Figure 3 shows a diagram of the single location test.



Site 1 – TX/RX

Fig. 3. Diagram of single location test

The two location test allows for the measurement of M2E latency between devices in two distinct locations. This test adds the capability to see potential effects of distance and propagation on the latency of devices. This test requires two laptops, two audio interfaces, one microcontroller, two communication devices, and two timecode generators. The time-code generators are locked to GPS clock signals, and add timestamps at both the transmit and receive locations. This allows for the audio signals to be aligned in post-processing in order to determine delay. The test is designed in much the same way as the single location test, as can be seen in Fig. 4a. The main difference is that both audio interfaces use a second input to record timecodes transmitted from the generator using the IRIG-B format and a cable modified to plug into the quarter-inch input jack on the interface.

The audio interface must have options to set the buffer size and USB streaming settings in order to minimize the possibility of audio glitches and buffer over/under runs. It also must have some signal that indicates whether the received audio is being clipped. Further, the audio interface must be able to simultaneously play and record audio via the MATLAB function *audioPlayerRecorder* in the Audio System Toolbox². Received audio from the interface is sampled at 48 KHz.

The measurement system presented is compatible with a variety of communication devices, but here the example measurements were performed with LMR radios. It should be

²Mathworks 2017, https://www.mathworks.com/help/audio/ref/audioplayerrecorder-class.html



(a) Lab diagram



(b) Field diagram

Fig. 4. Two location test diagrams

noted that the measurement system does not strictly require using MATLAB and could be implemented using other tools.

We now define the following terms: A trial is a single M2E latency test of an audio sample. A session consists of some number of consecutive trials. A test consists of some number of independent sessions. Due to common device constraints such as battery life and overheating, measurements must be segmented to minimize these factors. Thus it is necessary to identify the number of consecutive trials that can be performed without allowing these factors to impact results. By performing multiple sessions consisting of that many trials the amount of data that can be taken is not limited.

4.2 Delay Calculations and Audio Samples

Measuring delay over a communication system presents some inherent difficulties. If the speech is passed from input to output without significant distortion, then cross-correlation can be used without issue. However, in the case of a digital radio network the vocoder changes the waveform significantly, impeding the functionality of direct cross-correlation techniques.

To estimate delays the technique described in Ref. [3] is used. The audio signals are first downsampled to 8 kHz, then rectified and filtered to get their envelopes. Cross-correlation is then used on the envelopes of the transmitted and received signals to generate a delay value. Finally the delays are smoothed as described in Ref. [3]. The fixed delay algorithm described in Ref. [3] is used on a sliding window of audio. In particular, the audio signal is broken up into discrete, overlapping windows, and a single delay value is calculated for each window. This approach relies on the assumption that delay is fairly consistent within small enough windows of a single radio call. However, the sliding windows allow for the measurement of changes in delay across the call, due to events such as sporadically dropped packets.

For consistency, the same ten-second audio clip was used for all measurement tests. This clip consisted of four phonetically balanced sentences [4], with approximately 0.2-second gaps of silence between them. This clip can be found in Ref. [5], as well as with either the code or data associated with this paper.

The sliding delay estimate used four-second windows in two-second intervals. The size of the windows helped guarantee there was enough information within the windows for the algorithms to function properly. As an example, if the window size was too small, certain windows could be completely contained within some silent section of an audio clip, giving a cross-correlation with no significant peak, and making them indistinguishable from any other silent sections within the audio clip. This can cause significant errors within the delay calculations. A window size of four seconds with two-second slides on a ten-second audio clip results in four delay values for each radio call as shown in Fig. 5.

Due to the nature of the envelope cross-correlation technique at the core of the delay calculation, the integrity of the received signal is key to accurate delay measurements. If the received signal is significantly distorted, too noisy, or momentarily dropped, the reported delay values are not representative of the actual delay in the signal. As such it is key to control the quality of the signal as much as possible. Communicating via RF cable provides consistency for comparisons of radio communication technologies by removing external factors and real channel conditions from impacting delay values. For most systems the signal will need be attenuated to account for the losses that would normally occur with antennas and propagation. In initial tests, factors such as call duration and volume of the signal sent to the transmit device were seen to cause significant shifts in measured delay values. As such, the next two sections detail how each of these variables were controlled in



Fig. 5. Example audio envelope and windows. This is the envelope of the transmit audio used for all tests.

order to provide a fair platform on which test measurements may be compared regardless of the lab they were made in.

4.3 Maximizing Speech Quality

Receiving a high quality audio signal is important for reducing error in the delay measurement. If the audio quality is too low, then the peak in the cross-correlation (as described in Sec. 4.2) will be less pronounced and the algorithm can return erroneous results.

4.3.1 Ground Loop Effects

A ground loop is caused when there are multiple paths to ground in a single circuit. A ground loop can cause currents to flow in the ground path which can result in noise. During testing, noise was most pronounced when doing cabled testing in trunked mode. This is likely due to the large number of ground paths that were present.

One way to solve the ground loop problem is to use audio transformers in the audio path. A transformer magnetically couples the signal so there is no direct ground connection and no ground loop. Transformers have the benefit of being simple, passive devices and are easily inserted into the circuit.

Two different transformers were used. A simple 1:1 transformer was used between the receive radio and the audio interface. On the transmit side, the line output of the audio interface was fed into the microphone input of the transmit device. A transformer was inserted here to break a possible ground loop. The line output generates much higher levels than a microphone would, so a transformer that steps the signal down close to what a microphone would output is suggested. Due to this step down, it is necessary to consistently

use the line to microphone transformer in the volume setting procedure in the following section.

4.3.2 Audio Volume

The output volume used affects the audio quality. If the volume is too high, there will be clipping and if the volume is too low, then there will be significant noise. In general, audio interfaces have controls for the volume of the signal being sent from the playback outputs. In the context of the measurement system these control the volume of the speech being sent to the transmit device. In initial testing a difference in delay depending on the transmit volume was observed. Thus to provide a fair comparison for radio communication technologies, a procedure was needed to consistently control transmit volume.

Many communications devices such as LMR radios have audio processing features that aim to clean up low quality audio. These processes take time, so if low quality audio is sent to the device, the time for the device to transmit the audio increases. The transmit volume can affect audio quality in the following two ways: (1) If the transmit volume is set too low, the speech will be too quiet and it will be difficult to distinguish between the signal and the noise floor. (2) If the signal is too loud, the audio will be over driven and distorted. Knowing this, transmit volume levels were determined by audio quality.

Perceptual evaluation of speech quality (PESQ) is a common method of determining objective voice quality measurements [6]. Using an audio quality function, such as PESQ, the received audio can be directly compared to the transmitted audio to yield a value related to the quality of the received audio. With the ability to measure speech quality, the golden section search algorithm was used to identify the transmit volume associated with the maximum audio quality score. Note that the golden section search assumes a unimodal function, i.e. a function with a unique maximum. This assumption is reasonable in this context as the likely detractors for audio quality are having too low of a volume or too high of a volume. Thus it is fair to assume there is some continuous range of acceptable volume levels for which audio quality shall be fairly good. The golden section search will then return some value within this range. While this function is not necessarily strictly unimodal, any volume level that yields maximum audio quality is sufficient from the perspective of the communications devices, and thus sufficient for the consistent measurement of delay across devices, tests, and labs.

It should further be noted that most audio interfaces will also have a gain knob to control the received volume level. Received volume was not identified as significantly impacting delay values. However, having the received volume set too low can cause the signal to be lost, and set too high can cause the saved received audio to be severely clipped. As such, receive volume levels also need to be tracked. This proves difficult however due to the receive volume being controlled by a gain knob. For LMR devices a loudness units relative to full scale (LUFS) measurement can be performed using the P25 1011 test pattern as an input to the receive radio. As such, receive volume should be set to the highest level possible where clipping is consistently not observed. The following procedure describes how to set both the transmit and receive volumes for an audio interface:

1). Set V_{tx} to maximum volume

2). Adjust V_{rx} to max volume such that no clipping is observed

Play voice clip through radios and look for the clipping signal on the audio interface. It may take several runs through the voice clip to see clipping. If the clipping signal remains off, then turn the receive volume up. If there is clipping, the signal will indicate that the volume is too high and the receive volume should be lowered until no clipping is observed.

3). Record $V_{1011} \leftarrow V_{rx}$

To measure V_{1011} the transmit radio is disconnected from the receive radio and a 1011 pattern generator is connected to the receive radio (alternatively the transmit radio can be put in 1011 mode). The volume is measured with a MATLAB script. Because the 1011 pattern is fixed, the volume produced by it will always be the same.

4). Use golden ratio search to identify max audio quality

Reattach the transmit radio to the receive radio and run the golden ratio search. This will adjust the waveform scaling to test different output volume levels and determine which one has the best audio quality score. In order for the algorithm to function properly a good measure of the audio quality score needs to be taken. Due to natural variability in the communications devices, a large number of trials per volume level may be required. The variability of an audio quality score for a fixed volume can cause miscues within the golden section search algorithm, and result in inconsistent results. Fifty measurements were required for the example measurements presented later in this paper.

5). Set V_{tx} to the transmit volume associated with the max PESQ score

6). Repeat (2-5) until V_{1011} is not changing significantly.

4.4 Microcontroller

Prior to the inclusion of the microcontroller in the measurement system, initial testing showed that keying the radio into a transmit state for long periods of time tended to both increase the measured delay time and cause significant overheating in the radios. These tests involved having the radio transmit for upwards of 30 minutes, well beyond the intended design and use cases of the devices [7]. This implied the increased delay times that were measured were not indicative of the radio delay that would be experienced in normal use.

Thus the microcontroller was introduced in order to simulate more realistic radio use. The radios were keyed to transmit for 10-second periods which, while on the upper end of normal use, did not demonstrate the issues of the significantly longer transmit times. Note that the PTT timeout option of the radio must be set to be greater than whatever transmit length is used.

A Texas Instruments MSP430F5529 Launchpad board was used to interface between MATLAB and the radios. The MSP430F5529 Launchpad was used because it is a development board that integrates the target microcontroller with a programmer/debugger in a compact and economical package. The MSP430F5529 runs at a maximum clock speed of 25 MHz. There is 10 kB of on-chip RAM and 128 kB of on-chip flash. Programs are typically stored in the flash and run directly from it. The MSP430F5529 has 63 Input/Output (I/O) pins broken up into ports 1 through 8 and port J, which is multiplexed with the JTAG pins.

The software on the MSP connects over USB and presents itself as a virtual COM port. Text commands are sent to the MSP and parsed by the on board software. Text responses are sent back over the virtual COM port.

When the MATLAB code wants to activate the PTT, it sends a serial command to the MSP. To turn the PTT on, the MSP brings P8.2 (port 8 pin 2) high which turns on an optoisolator connected to the PTT wires of a headset connected to the radio. The schematic for the optoisolator is shown in Fig. 6.



Fig. 6. Circuit schematic for the optoisolator

Figure 7 shows the launchpad with the optoisolator circuit attached. The connections to power and I/O are accomplished through two dual row headers on the board. To control the optoisolator, P8.2 was used. The components were soldered to a perfboard to allow easy modifications.

4.5 Measurement Characterizations

The measurement system was characterized by replacing the communication devices with a cable that runs directly from the playback output of the audio interface to an input jack, as seen in Fig. 8, and performing M2E latency measurements. The main contributor to the latency of the measurement system is the audio interface, and as such this test is also referred to as a device characterization test.



Fig. 7. Microcontroller for controlling the PTT button on the radio



Laptop Running MATLAB Audio System Toolbox

Fig. 8. Diagram of single location system characterization

The two location measurement system was characterized in much the same way as the single location system, as seen in Fig. 9. Audio was sent out from the transmit site via a cable connected to RCA output 1 and connected directly to the receive site via the first quarter-inch input. Timecodes from the respective generators were recorded at both sites through the audio devices' second input.

The latency values reported from these tests are used to correct any latency results for a device test. In particular the characterization delays are subtracted from any device test, i.e. the two location characterization result would be subtracted from the results of a two location device test.



Fig. 9. Diagram of two location measurement characterization

4.6 Uncertainty Calculations

For a set of measurements x_1, \ldots, x_n , type A uncertainty is generally calculated as

$$u_{\overline{X}} = \frac{S_x}{\sqrt{n}},\tag{1}$$

where \overline{X} is the mean value of the measurements, S_X is the sample standard deviation, and *n* is the number of trials. However, this technique is only valid for data that is probabilistically independent [8]. As the delay calculations of Sec. 4.2 return delay values corresponding to overlapping windows, the delay measurements returned for a single trial are clearly dependent. As a way around this, the delay values corresponding with one trial were averaged into a single measurement value in the hopes that this would yield independent data across trials. However, even after this was done it could be shown that the data within sessions was autocorrelated.

Thus it was determined that each session of a test needed to be treated separately, and that autocorrelation would be removed from the data by thinning it. Thinning was accomplished for a test by using only every *m*th measurement, where *m* is chosen to be as small

as possible while still eliminating statistically significant autocorrelation in all thinned sessions of that test. As an example consider a set of ten data points, x_1, \ldots, x_{10} thinned by using every third measurement. Then only data points x_1, x_4, x_7, x_{10} would be used for all further analysis and reporting of results. By increasing the time between measurements that are used for analysis, significant autocorrelation can be eliminated. Significant autocorrelation was determined as described in Ref. [8].

The limited resolution of the measurement also affected the uncertainty. The delay calculations require the audio be downsampled to 8 kHz, corresponding with a measurement resolution of $\delta = 0.125$ ms. This type B uncertainty was incorporated as described in Ref. [9].

The combined standard uncertainty of a measurement characterization test is then reported as

$$u_c(y) = \sqrt{\frac{\delta^2}{12} + \frac{1}{N^2} \sum_{k=1}^N u_k^2}$$
(2)

where N is the number of measurement sessions for the test and u_k is the uncertainty associated with the thinned data of session k. In particular, u_k is calculated via Eq. (1) using the thinned data of session k.

The combined standard uncertainty of a device M2E latency test is reported as

$$u_c(y) = \sqrt{u_{\text{character}}^2 + \frac{\delta^2}{12} + \frac{1}{N^2} \sum_{k=1}^N u_k^2}$$
(3)

where again N is the number of measurement sessions for the test and u_k is the uncertainty associated with the thinned data of session k. Here $u_{character}$ represents the standard uncertainty associated with the relevant measurement characterization. This term captures the uncertainty associated with correcting measurements for the inherent delay of the measurement system.

It should be noted that the resolution of the received timecodes can also affect measurement uncertainty. However, it is unlikely that this would significantly contribute to the overall uncertainty. In the case of the example measurements presented in the next section, the timecode generators are accurate up to 1 μ s and are sampled at 48 KHz. Even after accounting for the limited resolution of the sampling of both the transmit and receive timecode generators, the change in uncertainty was insignificant.

Uncertainty for each test is reported with coverage factor, k, associated with a 95% confidence level. Thus latency values are reported as $\overline{X} \pm U$, where \overline{X} is the mean of all the thinned sessions for a particular test and $U = ku_c$ is the expanded uncertainty. Thus the unknown value of the measurand is believed to lie in the interval defined by U with a level of confidence of approximately 95%. Coverage factor k will be specified for each test.

5. Example Delay Measurements

A series of example measurements were performed to characterize both UHF and VHF handheld radios in direct and trunked mode. All lab tests were performed communicating via cables with field tests performed over the air.

Behringer UMC 204HD audio interfaces recorded and played back audio data. The buffer size was set to *512* samples and the USB Streaming set to *Standard* to avoid buffer over/under runs and audio glitches respectively. For device tests a TI MSP430 micro-controller was used to toggle the PTT switch for the radios. For two location tests, ESE ES-185E GPS Master Clock timecode generators were used. The transformer between the line input on the audio interface and the transmit device was a Jensen JT-DB-E, with a gain of -22 dB. The transformer between the receive radio and the audio interface was a Jensen JT-11P-1.

5.1 Additional Measurement Information

For both the single and two location measurements data was collected through a sequence of measurement sessions. Data collection has to be segmented in this way due to device constraints such as battery life and overheating. For each trial, four delay measurements were returned from the *sliding_delay_estimates* function described in Ref. [3]. For all tests presented here sessions consisted of 100 consecutive trials, with a 3.5 second pause between each trial. Data was collected via four measurement sessions for all tests.

Two sets of two location measurements were performed: first in a lab test with both the transmit and receive devices communicating via cabled RF to verify that the measurement system behaved the same as the single location system, and then in a field test with the receive device in the lab and the transmit device 6.99 km away. The measurement sites for the field test are shown in Fig. 10.

It should be noted that propagation is expected to have minimal impact on overal M2E latency. RF travels at the speed of light which in air is approximately 3×10^8 m/s. The maximum distance for the two location test was conducted with the LMR approximately 7 km from the repeater antenna located on top of the Wing 2 roof of Building 1 at the NIST Boulder campus. At a distance of 7 km (7,000 m), the propagation delay is calculated to be approximately 23 microseconds. With M2E latency values in the multiple millisecond range, the RF propagation delay is practically negligible. Further, it was found that the propagation delay was smaller than the uncertainty in the measurements, and thus should have no effect on measurement results.

In order for the radios to communicate with the audio interface, external 3-wire microphone/earphone/PTT button devices were modified, as seen in Fig. 11. For the transmission radio, the PTT switch was replaced with a connection to the output of the optoisolator on the microcontroller, which allowed the PTT to be controlled by MATLAB. The transmit microphone was also replaced with an RCA jack and connected to the playback output on the audio interface, so that audio could be sent directly from the computer to the audio



Fig. 10. Two location field measurement sites. Sites are 6.99 km apart.

interface and finally to the radio. On the receive side, the earphone was replaced with an XLR connector that was then connected to an input on the audio interface.



Fig. 11. Adapted LMR headsets

Longer delays tended to occur during the over the air field tests as compared to the cabled RF lab tests. In over the air tests the radio devices are subjected to real channel conditions which may cause the receiver to require more time to process the received signal. Effects such as noise and multiple signal paths between the transmitter and receiver both can cause additional processing in the receiver.

Any data that showed autocorrelation was thinned by the minimum amount that eliminated significant autocorrelation for all sessions of a particular test. The degree to which tests were thinned are reported in Table 2 as well as with the measurement results in the following sections. All reported values are calculated using the thinned data. It should be noted that all M2E latency values will be presented as $\overline{X} \pm ku$ where ku is the expanded uncertainty associated with a level of confidence of 95%.

Table 2. Test Thinning Information. These values represent the degree to which the data for each test had to be thinned in order to eliminate significant autocorrelation. A value of 1 means that every data point was used, and hence the data was not thinned, 2 means that every other data point was used, etc...

	Single Location Lab	Two Location Lab	Two Location Field
Measurement	3	4	4
Characterization			
UHF-Direct	4	3	5
UHF-Trunked	4	5	4
VHF-Direct	7	5	4
VHF-Trunked	3	5	1

5.2 Measurement Characterization

The single location measurement system characterization results are shown in Fig. 12a. Only three results are returned: 21.750, 21.875, and 22.000 ms. These values differ by 0.125 ms steps. The value of 0.125 ms characterizes the resolution of the measurement. As the expected sampling rate of the *ITS_delay_est* function [3] is 8000 Hz, measurements are forced to be discretized in 1/8000 = 0.125 ms intervals. Thus the measurement system contributes a very consistent delay. After thinning the data by using every third measurement to remove autocorrelation the average system delay is 21.85 ± 0.07 ms with a coverage factor of k = 1.96.

The value of 0.07 ms characterizes the single location test measurement uncertainty as described in Sec. 4.6.

The two location measurement system has the same resolution of the measurement system as the single location, and the same three values were measured: 21.750,21.875, and 22.000 ms, as seen in Fig. 12b. After thinning the data by using every fourth measurement, the two location system measurement system latency is 21.85 ± 0.07 ms, with a coverage factor of k = 1.96. This result lines up very well with the single location measurement.

5.3 UHF Direct Mode

The single location UHF direct mode M2E latency measurements are shown in Fig. 13a. After thinning the data by using every fourth measurement to remove autocorrelation the



Fig. 12. Measurement characterization results. Vertical dotted black lines denote separate data collection sessions.

system corrected latency is 201.4 ± 0.4 ms with a coverage factor of k = 1.98. Again the uncertainty of 0.4 ms was calculated as described in Sec. 4.6.

The two location lab UHF direct mode M2E latency measurement results are shown in Fig. 13b. The data was thinned by using every third measurement and the corresponding system corrected M2E latency is 201.2 ± 0.3 ms with a coverage factor of k = 1.97. Thus the single and two location measurement systems return very similar results with significant overlap in their confidence intervals, and are thus consistent.

The two location field test measurement results are shown in Fig. 13c. The data was thinned by using every fifth measurement and the corresponding system corrected M2E latency is 201.8 ± 0.4 ms, with a coverage factor of k = 1.99.

There is overlap between the confidence intervals for all three tests, suggesting the tests are consistent.

5.4 UHF Trunked Mode

The single location UHF trunked mode M2E latency measurement results are shown in Fig. 14a. The data was thinned by using every fourth measurement and the corresponding system corrected latency is 415.8 ± 2.8 ms. As can be seen in Fig. 14a, the delays in trunked mode are subject to much more variability than in direct mode. Trunked mode communications is a more complicated process with more steps in the communications channel than direct mode, and it is very likely these extra steps all add a degree of variability to the overall delay.

The two location lab UHF trunked mode M2E latency measurement results are shown in Fig. 14b. After thinning the data by using every fifth measurement to remove autocorrelation the system corrected M2E latency is 413.1 ± 3.3 ms, with a coverage factor of k = 2.00. There is again overlap between the single and two location tests, suggesting they return consistent results.

The two location field test measurement results are shown in Fig. 14c. The data was thinned by using every fifth measurement, and the corresponding system corrected M2E latency is 417.0 ± 2.9 ms, with a coverage factor of 1.99.

Due to the overlap of all respective confidence intervals, all three tests returned consistent results for UHF trunked mode.

5.5 VHF Direct Mode

The single location VHF direct mode test measurement results are shown in Fig. 15a. The data was thinned by using every seventh measurement, and the system corrected M2E latency is 201.7 ± 0.5 ms, with a coverage factor of 2.00.

The two location lab test measurement results are shown in Fig. 15b. The data was thinned by using every fifth measurement and the system corrected M2E latency was found to be 201.6 ± 0.4 ms, with a coverage factor of 1.99. Thus both the lab tests are consistent with each other.



Fig. 13. Raw UHF direct mode measurements. Vertical dotted black lines denote separate data collection sessions.



Fig. 14. Raw UHF trunked mode measurements. Vertical dotted black lines denote separate data collection sessions.

The two location field test measurement results are shown in Fig. 15c. The data was thinned by using every fourth measurement and the system corrected M2E latency was found to be 202.4 ± 0.4 ms, with a coverage factor of 1.98.

Both of the lab tests have very close results, with significant overlap of their confidence intervals and are thus consistent. The single location test and the two location field test also have overlap between their confidence intervals and are thus consistent, however the two location lab and two location field tests do not. The intervals are still very close to overlapping, and the discrepancies are likely due to factors of real channel conditions in over the air communications that affect the communications channel.

5.6 VHF Trunked Mode

Single location VHF trunked mode M2E latency measurement results are shown in Fig. 16a. After thinning the data by using every third measurement, the system corrected M2E latency is 403.9 ± 1.8 ms, with a coverage factor of 1.98.

The two location lab test measurement results are shown in Fig. 16b. After thinning by using every fifth measurement, the corrected M2E latency was found to be 403.3 ± 2.8 ms, with a coverage factor of 1.99. Thus again the lab tests can be shown to be consistent.

The two location field test measurement results are shown in Fig. 16c. The data for this test was not autocorrelated, and was thus not thinned. The corrected M2E latency was found to be 405.3 ± 1.2 ms, with a coverage factor of 1.97.

Like the UHF trunked tests, all VHF trunked tests returned consistent results.

6. Conclusion

Due to the consistency between the lab measurements for all tests, it is safe to conclude that the single and two location tests are capturing M2E latency for devices in the same way. The slight discrepancies in the results of some of the field tests speak more to the measurement systems ability to pick up small variations in delay caused by the real channel conditions of the over the air test. The single location system is ideal for baselining device M2E latency. Due to the sensitivity of the measurement system to signal conditions and noise from the devices, it is best to use a cabled RF connection in a lab setting rather than using over the air communication when comparing radio communications technologies. This helps to provide consistency across tests so that fair comparisons can be made. The two location system further allows for the measurement of propagation effects or cross radio site communication delay values.

The uncertainty in measurement contributed by the measurement system is largely insignificant compared to the inherent variability of M2E delay within the devices tested. It is expected that this would be consistent across the majority of communications devices. Both measurement systems presented utilize an audio in/audio out approach to delay measurement that maintains device independence.



Fig. 15. Raw VHF direct mode measurements. Vertical dotted black lines denote separate data collection sessions.



Fig. 16. Raw VHF trunked mode measurements. Vertical dotted black lines denote separate data collection sessions.

References

- [1] 3GPP (2017) Mission Critical Push to Talk (MCPTT). 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.179. Version 16.0.0 URL https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx? specificationId=623.
- [2] TIA (2016) Digital C4FM/CQPSK Transceiver Measurement Methods. Telecommunications Industry Association (TIA), Standard 102.CAAA-E.
- [3] Voran SD (2004) A bottom-up algorithm for estimating time-varying delays in coded speech. *Proceedings of the 3rd International Conference on Measurement of Speech and Audio Quality in Networks*.
- [4] (1969) IEEE Recommended Practice for Speech Quality Measurements. *IEEE No 297-1969* :1–24https://doi.org/10.1109/IEEESTD.1969.7405210
- [5] ITU-T (1998) Coded-speech database. International Telecommunications Union Telecommunication Standardization Sector (ITU-T), Recommendation P Suppl. 23. URL http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=4415&lang=en.
- [6] ITU (2005) Telephone Transmission Quality, Telephone Installations, Local Line Networks. International Telecommunication Union (ITU), Recommendation P.862. Amendment 2 URL https://www.itu.int/rec/T-REC-P.862.
- [7] Sharp DS, Cackov N, Laskovic N, Shao Q, Trajkovic L (2004) Analysis of public safety traffic on trunked land mobile radio systems. *IEEE Journal on Selected Areas in Communications* 22(7):1197–1205.
- [8] Zhang NF (2006) Calculation of the uncertainty of the mean of autocorrelated measurements. *Metrologia* 43(4):S276. URL http://stacks.iop.org/0026-1394/43/i=4/a=S15.
- [9] Lira IH, Wöger W (1997) The evaluation of standard uncertainty in the presence of limited resolution of indicating devices. *Measurement Science and Technology* 8(4):441. URL http://stacks.iop.org/0957-0233/8/i=4/a=012.