NIST Technical Note 2166

Mission Critical Voice Start-of-Word Correction for Access Delay Measurement System

William Magrogan
Jaden Pieper
Zainab Soetan

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.2166

National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Technical Note 2166

Mission Critical Voice Start-of-Word Correction for Access Delay Measurement System

William Magrogan
Jaden Pieper
Zainab Soetan
Public Safety Communications Research Division
Communications Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.2166

June 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Technical Note 2166
Natl. Inst. Stand. Technol. Tech. Note 2166, 29 pages (June 2021)
CODEN: NTNOEF

This publication is available free of charge from:
https://doi.org/10.6028/NIST.TN.2166

Acknowledgments

The authors would like to thank Steve Voran of the Institute for Telecommunications Sciences for his insight in suggesting a start-of-word correction, his help in selecting keywords for access delay tests, and for all of his discussions and assistance along the way. The authors would like to thank Jesse Frey for his help collecting data for the start-of-word correction. Finally, the authors would like to acknowledge the entire mission critical voice quality of experience measurements team.

Abstract

NIST's Public Safety Communications Research (PSCR) division previously developed an access time measurement method [1] for push-to-talk (PTT) communications systems. This paper introduces a correction term, called the start-of-word correction, to address a systemic error in the original measurement method. This error can be caused by extraneous audio prior to the start of a Modified Rhyme Test (MRT) keyword in the test audio clips that were developed for the access time measurement. Example measurements were performed using the start-of-word correction to quantify its affect on access time measurements.

Key words

Access delay; Access time; Articulation Band Correlation Modified Rhyme Tests (ABC-MRT); ABC-MRT16; Land mobile radio (LMR); Long term evolution (LTE); Mission critical voice (MCV); Modified Rhyme Test (MRT); Project 25 (P25); Public safety; Push-to-talk (PTT); Quality of experience (QoE); Trunked mode.

Table of Contents

List of Tables

List of Figures

Acronyms

3GPP 3rd Generation Partnership Project. 3, 4

ABC-MRT Articulation Band Correlation Modified Rhyme Test. 1, 12

E2E end-to-end. 1

LMR land mobile radio. 12

LTE Long Term Evolution. 12, 14

M2E mouth-to-ear. 1, 6

MCPTT Mission Critical Push-to-Talk. 4

MRT Modified Rhyme Test. i, 1, 2, 6, 7, 9, 12, 13, 15

P25 Project 25. 14

PSCR Public Safety Communications Research. i, 3

PTT push-to-talk. i, 1–9, 11–14

QoE quality of experience. 3

QoS quality of service. 3, 4

RMSE root mean square error. 12

SNR signal-to-noise ratio. 13

SUT system under test. 1–15

TIA Telecommunications Industry Association. 3

UE user equipment. 4

WUT word under test. 1, 3–6, 9–15, 19

Symbols

α Intelligibility scaling factor. 1, 3–6, 8–11

I_0 Asymptotic intelligibility. 2, 3, 6, 11

λ Logistic parameter, intelligibility curve steepness. 2–6, 9–11

λ_{s^*} The intelligibility curve steepness parameter for the SUT s^* for a WUT. 4, 5

P_1 First utterance of MRT keyword. 2, 12

P_2 Second utterance of MRT keyword. 2, 12

S The set of all possible SUTs that perform no audio buffering prior to PTT. 4–6

s^* The SUT that minimizes access delay for a WUT. 4–6

T Time preceding P_1 and P_2 in audio clips. 2

t_0 Logistic parameter, intelligibility curve midpoint. 2–5, 10

t_{0,s^*} The intelligibility curve steepness parameter for the SUT s^* for a WUT. 4, 5

τ Access delay, function of α. 3–5, 7–9

τ_{P} Access delay inherent to audio clip, used as measurement correction. 6, 9

τ_{S} Access delay of an audio clip through a SUT. 8, 9

□ End of proof. 4, 5

1. Introduction

This paper introduces a measurement correction for the access delay measurement system developed in Ref. [1] and further refined in Ref. [2]. In particular, it addresses a systemic error that can be introduced from the Modified Rhyme Test (MRT) keyword audio clip structure. The primary source of this systemic error is driven by audio immaterial to the intelligibility of the MRT keyword preceding the actual keyword in the audio clips used in an access delay measurement.

The correction accounts for such audio by measuring the access delay of keywords through a hardware approximation of the minimal access delay push-to-talk (PTT) communications system: the PTT gate introduced in Ref. [1]. For a given set of MRT keywords, the correction is implemented via post-processing by incorporating PTT gate results with results through any system under test (SUT). Thus, any previously collected data can in theory be corrected with the new addition. In practice, however, it was observed that the new curve fitting procedure used for the correction was more sensitive to the behavior of specific keywords, so previously collected data may not be suitable for the new correction.

Intelligibility in this paper refers specifically to MRT intelligibility [3] and intelligibility estimates were calculated using the objective speech intelligibility algorithm Articulation Band Correlation Modified Rhyme Test (ABC-MRT). In this paper, ABC-MRT refers to the version ABC-MRT16 [4]. Further, MRT keywords refer to audio extracted from the modified rhyme test audio library [5].

2. Background

2.1 Access Time Measurements

In Ref. [1] we introduced the definition and a measurement system for access delay. Along with mouth-to-ear (M2E) latency, access delay is a component of end-to-end (E2E) access time. Access delay is defined as the minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the first word of the message has an average intelligibility that is no lower than $\alpha \cdot I_0$, where $0 < \alpha < 1$ is a scaling parameter, and I_0 is the asymptotic intelligibility of that word through the communications system. Here, asymptotic intelligibility describes the intelligibility of a given word under test (WUT) through a SUT that is fully operational in a transmit mode. Thus, access delay measurements are dependent on both a given WUT and the specific SUT. It should be noted that WUT refers to a particular combination of an MRT keyword and individual talker, so under this definition, the same word spoken by two different individuals counts as two independent WUTs.

The definition of access delay relies on an accurate notion of the time between when PTT is pressed and the first word of the message being spoken. The measurement system uses predefined audio clips that are built from an existing database of MRT keyword recordings [5]. These MRT keywords were extracted from full MRT phrases that include

both the carrier phrase and the keyword. An example phrase is "Please select the word *west*."

Determining exactly when the MRT keyword is spoken within this whole phrase can be difficult. In general, there is co-articulation between the end of the carrier phrase and the beginning of the MRT keyword. This means there is no clear boundary between when the carrier phrase is fully spoken and the MRT keyword starts to be spoken. When keywords were originally extracted, cutpoints were chosen conservatively, opting to include additional audio at the front of keywords rather than risk cutting out important information. Due to this, there is potential for extraneous audio to be present in the extracted MRT keyword audio clips that are used to design test signals for access delay measurements. If unaccounted for, the length of this peripheral audio impacts access delay measurements and can cause reported access delays to be under-estimates of the true value. This motivates the need for a start-of-word correction in the access time measurement system.

2.2 Measurement Overview

Audio clips for an access delay measurement are structured as follows. One first selects an MRT keyword recording to use for the measurement. One then chooses a value of T such that T is almost certainly larger than the access delay of the SUT. T should be chosen conservatively, and in our previous measurements we have found $T = 2.5$ s to generally be a good selection [2]. The audio clip begins with T seconds of silence, before the first play-through of the MRT keyword, denoted P_1. T seconds of filler speech are then played, followed by a second utterance of the MRT keyword, called P_2.

Data collection for an access delay measurement involves varying when PTT is pressed in relation to when the audio clip is played into a transmit device. A test starts with PTT occurring right at the end of P_1 being spoken. This guarantees that P_1 is not transmitted, and that the corresponding receive audio has an intelligibility of zero for P_1. Further, the selection of T attempts to guarantee that by the time P_2 is transmitted, the SUT is fully set up, and P_2 is transmitted and received at the SUT's asymptotic intelligibility for that keyword. The test proceeds by pressing PTT earlier and earlier. As PTT time becomes earlier, there is more and more time between PTT being pressed and when P_1 is sent through the transmit device. Eventually, P_1 is being sent through the transmit device with enough time preceding it to achieve the same intelligibility of P_2, the asymptotic intelligibility.

2.3 Original Curve Fitting Procedure

Once data is collected, it is processed to make access delay measurements. This primarily involves fitting a logistic curve to the collected P_1 intelligibility values. We define an intelligibility function as

$$I(t) = \frac{I_0}{1 + e^{(t-t_0)/\lambda}} \tag{1}$$

where t is the time between PTT being pressed and P_1 going into the transmit device. This function relies on three parameters that are determined from the data: $I_0 \in [0,1]$ is the

asymptotic intelligibility of the WUT through the SUT, $t_0 \in \mathbb{R}$ is the time associated with the midpoint of the intelligibility curve, and $\lambda < 0$ is, up to a constant factor, the reciprocal of the slope at $I(t_0)$.

This intelligibility function directly describes how intelligible the WUT is given there are t seconds between PTT being pressed and the WUT going into the transmit device. For example, $I(0)$ is the intelligibility of the WUT when it goes into the transmit device immediately upon PTT being pressed. $I(t)$ for $t < 0$, describes the intelligibility of the WUT when PTT is pressed after the WUT has started to go through the transmit device. When $t < 0$ and the magnitude of t is less than the length of the WUT, the WUT was said in its entirety before PTT was pressed, and intelligibility will be essentially 0. Finally, $I(t)$ when $t > 0$ describes the standard use case, where a user waits some amount of time after pressing PTT before speaking into a device.

The curve parameters have the following interpretations: I_0 describes the intelligibility of the WUT through the SUT when the system is fully turned on and ready to transmit at normal operating condition, t_0 describes the amount of time required between pressing PTT and speaking the WUT to achieve an intelligibility of $I_0/2$, and λ relates to the steepness of the intelligibility transition. In other words, λ relates to the width of the region where intelligibility transitions from 0 to I_0.

If you set $I(t) = \alpha \cdot I_0$ in Eq. (1) and solve for t, the result will be a function of α and we call it $\tau(\alpha)$. This new function describes access delay:

$$\tau(\alpha) = \lambda \ln\left(\frac{1-\alpha}{\alpha}\right) + t_0 \tag{2}$$

In general, to robustly measure access delay for a SUT, one must capture data across a variety of keywords and talkers. The curve fitting procedure in Ref. [1] opted to fit a single curve to an aggregate of intelligibility data from multiple measurement sessions of different talker-word combinations. This provides the benefit of making access delay measurements less sensitive to erratic behavior in intelligibility estimations for different keywords. The start-of-word correction is not compatible with a single curve being fit to an aggregate of all WUT data, so access delay measurements relying on the correction will be more sensitive to word selection, as will be discussed further in Sec. 5.1.

2.4 Access Time in Standards

It is important to emphasize that access delay and end-to-end access time discussed in the context of Public Safety Communications Research (PSCR)'s quality of experience (QoE) measurements are distinct from similar definitions defined by standards bodies. In particular, access delay relies solely on speech intelligibility to determine whether or not access was granted. In contrast, access time standards, such as the ones defined by the Telecommunications Industry Association (TIA) and the 3rd Generation Partnership Project (3GPP), rely on quality of service (QoS) and technology-specific metrics.

In particular, the TIA defines access time as "the time required from the initiation of a user push-to-talk until a traffic channel is granted, and transmission on that channel has

begun. This measurement includes the time for both the subscriber equipment and the RF [radio frequency] subsystem equipment" [6]. This definition focuses completely on the output power of the transmitting radio.

Similarly, the 3GPP defines Mission Critical Push-to-Talk (MCPTT) access time as "the time between when an MCPTT User requests to speak (normally by pressing the MCPTT control on the MCPTT user equipment (UE)) and when this user gets a signal to start speaking" [7]. They also define end-to-end access time as "the time between when an MCPTT User requests to speak (normally by pressing the MCPTT control on the MCPTT UE) and when this user gets a signal to start speaking, including call establishment (if applicable), and possibly acknowledgment from first receiving user before voice can be transmitted" [7]. Again, both of these definitions are QoS based; they describe access time in terms of system signaling and device acknowledgment. While the end-to-end access time definition accounts for receiving devices acknowledging the channel grant, they do not account for the crux of a user experience: how long must the talker wait in order to ensure that intelligible voice is received by another user.

3. Constraining Curve Parameters

Before formally defining the start-of-word correction, we aim to constrain the possible curve fit parameters λ and t_0 for a given WUT. We do this by considering a theoretical SUT that minimizes access delay for a given WUT.

3.1 A Minimal Access Delay Communications System

Let S be the set of all possible SUTs that perform no audio buffering prior to PTT and let W be the set of all WUT. We define $s^* \in S$ as the SUT that minimizes access delay for a given $w \in W$ over the interval $\alpha \in [1/2, 1)$. In particular for a given $w \in W$

$$\tau_{s^*}(\alpha) := \min_{s \in S} \tau_S(\alpha), \quad \forall \alpha \in [1/2, 1) \tag{3}$$

This means that for any $s \in S$, $\tau_{s^*}(\alpha) \leq \tau_s(\alpha) \ \ \forall \alpha \in [1/2, 1)$. With eq. (2) we define λ_{s^*} and t_{0,s^*} such that

$$\tau_{s^*}(\alpha) = \lambda_{s^*} \ln\left(\frac{1-\alpha}{\alpha}\right) + t_{0,s^*}$$

Lemma 3.1 :

$$t_{0,s^*} \leq t_{0,s}, \quad \forall s \in S \tag{4}$$

Proof. Let $\alpha = 1/2$, then $\tau_s(1/2) = \lambda_s \ln(1) + t_{0,s} = t_{0,s}$. From eq. (3) it follows that $t_{0,s^*} \leq t_{0,s} \ \forall s \in S$. □

Lemma 3.2 :

$$\lambda_{s^*} \geq \lambda_s, \quad \forall s \in S \tag{5}$$

Proof. Let $s' \in S$ be defined as

$$\tau_{s'}(\alpha) = \lambda_{s'} \ln\left(\frac{1-\alpha}{\alpha}\right) + t_{0,s^*}$$

Note that $\tau_{s'}(\alpha) \leq \lambda_{s'} \ln\left(\frac{1-\alpha}{\alpha}\right) + t_{0,s}$ for any $s \in S$ from lemma 3.1. From eq. (3) it follows that

$$\begin{aligned} & \tau_{s^*}(\alpha) \leq \tau_{s'}(\alpha) \\ \implies & \lambda_{s^*} \ln\left(\frac{1-\alpha}{\alpha}\right) + t_{0,s^*} \leq \lambda_{s'} \ln\left(\frac{1-\alpha}{\alpha}\right) + t_{0,s^*} \\ \implies & \lambda_{s^*} \ln\left(\frac{1-\alpha}{\alpha}\right) \leq \lambda_{s'} \ln\left(\frac{1-\alpha}{\alpha}\right) \end{aligned}$$

Note that $\ln\left(\frac{1-\alpha}{\alpha}\right) < 0 \; \forall \alpha \in (1/2, 1)$, so

$$\lambda_{s^*} \geq \lambda_{s'}$$

□

Theorem 3.3 : *For a given WUT, w, let S be the set of all possible SUTs that perform no audio buffering prior to PTT, and let s^* be the SUT in S that minimizes access delay over $\alpha \in [1/2, 1)$ for w, such that for any $s \in S$, $\tau_{s^*}(\alpha) \leq \tau_s(\alpha) \quad \forall \alpha \in [1/2, 1)$. Then the access delay curve parameters λ_{s^*} and t_{0,s^*} for s^*, provide upper and lower bounds respectively for possible curve parameters of w through any SUT, $s \in S$.*

Proof. See lemmas 3.1 and 3.2. □

Theorem 3.3 considers only SUTs that do not buffer audio prior to PTT being pressed. We claim that a SUT s that performs audio buffering prior to PTT being pressed could satisfy $t_{0,s} < t_{0,s^*}$, but could not have a value of λ_s that satisfies $\lambda_s > \lambda_{s^*}$. This is because audio buffering can only cause a constant offset in access delays and would not change the intelligibility transition; that is, it would only affect t_0 and not λ. In other words, for any SUT with a finite audio buffer, one could design an access delay measurement test that captures the intelligibility transition. Thus, we claim that when expanded to the set of SUTs that can contain audio buffering, lemma 3.1 does not hold, but lemma 3.2 does.

Theorem 3.3 only requires that s^* minimize access delay over the region $\alpha \in [1/2, 1)$. This domain restriction is caused by a limitation of the model selection[1], namely, there

[1] Any $\mathbf{C}^1$ continuous model with asymptotic limits at 0 and 1 will have this same limitation.

exists no single SUT that minimizes access delay over all of $\alpha \in (0,1)$. In particular, the SUT that minimizes access delay over $(0,1/2]$, does not minimize access delay over $[1/2,1)$, and vice versa. In reality, this is not a pertinent constraint and is only a limitation of the chosen model. Thus, we chose to focus on the interesting intelligibility domain of $\alpha \in [1/2,1)$, as access delay for higher intelligibility is more practical and useful.

3.2 Real World Approximation

Ref. [1] introduced a simple PTT communications system to help characterize the access delay measurement system, called the PTT gate. Essentially, the PTT gate is a simple switch placed between two loop back cables. It blocks audio going through until PTT is pressed and again blocks when PTT is released. It was designed to have minimal yet consistent access and M2E latency values. The amount of M2E latency introduced by the PTT gate is smaller than the resolution of our measurement, so when measured we see effectively 0 ms of latency. The specifications for the PTT gate say the maximum turn-on time is 2 ms. However, in lab testing the average observed value and 95% confidence interval was 93.708 ± 0.005 µs. Thus, in practice, the PTT gate introduces no measurable M2E latency or access delay as a SUT.

Due to the lack of access delay introduced by the PTT gate, denoted as $\mathrm{P} \in S$, we argue that P is a suitable approximation of the minimal access delay SUT, s^*, defined in Sec. 3.1. While the PTT gate can potentially introduce extremely small amounts of access delay, these observed turn-on times of the device are an order of magnitude smaller than the length of phonemes, and would likely have no impact on the intelligibility scores seen in an access delay measurement. We claim that the access delay parameters from the PTT gate, λ_{P}, $t_{0,\mathrm{P}}$, are reasonable approximations on the bounds of possible curve parameters for any WUT introduced in theorem 3.3.

Similarly, because the PTT gate introduces minimal access delay, any measurement of access delay with the PTT gate as the SUT will always return negative access delay. This is because

$$I(0) \approx I(t) \approx I_0, \quad \forall t > 0.$$

Intuitively, the above can be described by the fact that adding additional silence before speaking a WUT does not add additional speech information, and thus does not increase the intelligibility of the first word in a message.

This all implies that access delay values for the PTT gate can be interpreted as being descriptive of critical points within MRT keyword audio. In other words, $\tau_{\mathrm{P}}(\alpha)$ for a WUT through the PTT gate describes how much of the beginning of the WUT audio can be removed in order to achieve an intelligibility of $\alpha \cdot I_0$.

4. Technical Approach

4.1 Motivation

To help solidify the need for a start-of-word correction, we present the following example. All the numbers presented here are constructed purely to provide a motivating example for the utility of the start-of-word correction and do not represent reality. Consider measuring access delay of the MRT keyword "hook" through some SUT. For the purposes of this example, fix $\alpha = 0.9$.

Let us first consider the structure of this "hook" audio clip. An imagined structure for this example keyword audio is shown in Fig. 1, where the audio had 50 ms of leading silence. This silence at the beginning of the audio clip for "hook" has no impact on its intelligibility. The word "hook" also has a long "h" sound at its beginning. Let it take 200 ms to fully articulate this "h" sound. However, let us assume that only the last 100 ms of the long "h" sound are necessary to distinguish it from similar sounding words, in the case of MRT tests, "took", "cook", "look", "shook", and "book". In other words, the leading 100 ms of the actual speech containing "hook" can be muted and the full word would still be understood 90% of the time. This means that the first 150 ms of this audio clip can be muted and the keyword would still be intelligible 90% of the time.

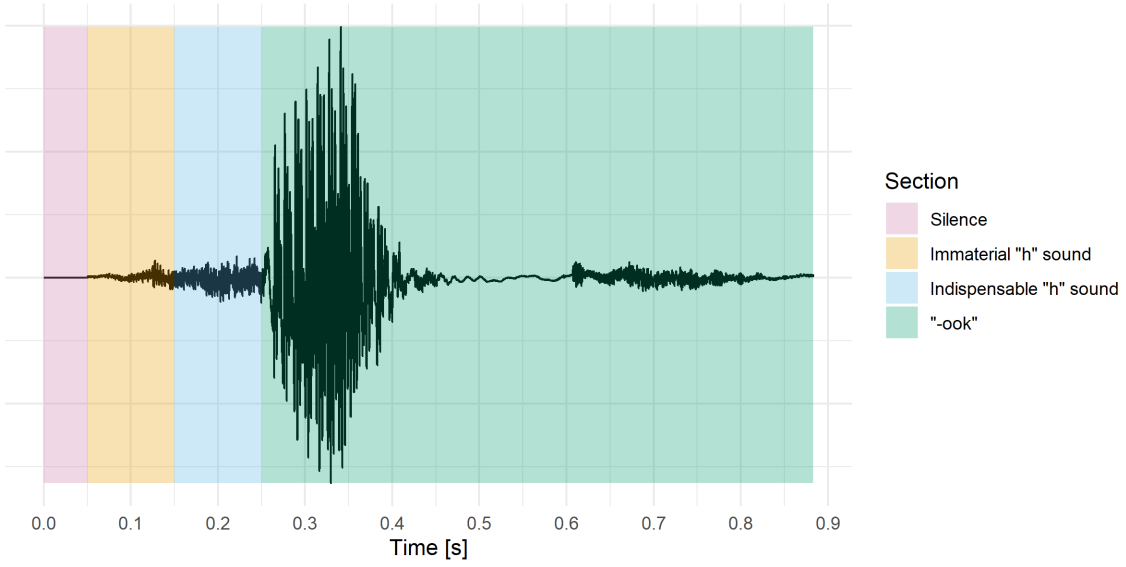

Fig. 1. Structure of the example audio for "hook". This keyword comes from the MRT list containing "took", "cook", "look", "hook", "shook", and "book", so the "-ook" sound does not contribute any intelligibility information in this example.

If access delay was measured on this "hook" clip through the PTT gate, then this result would be observed, where $\tau_P(90\%) = -150$ ms. Access delay measurements with the PTT gate allow us to characterize the critical intelligibility points within an audio clip containing an MRT keyword.

Consider next measuring access delay on this "hook" clip through some arbitrary SUT, $\tau_S(\alpha)$. Let us say that access delay is measured to be 0 ms. Without a correction term, this result would lead someone to incorrectly think there is no access delay for this keyword in the SUT. However, as this example makes clear, there is 150 ms of audio at the beginning of this clip that has no impact on the 90% intelligibility of "hook". This is 150 ms that is hidden in the 0 ms access delay result from the SUT. Explicitly, if the SUT actually had no access delay it would return an access delay of -150 ms, just as the PTT gate does. Intelligibility curves for these example measurements using both the PTT gate and SUT are shown in Fig. 2.

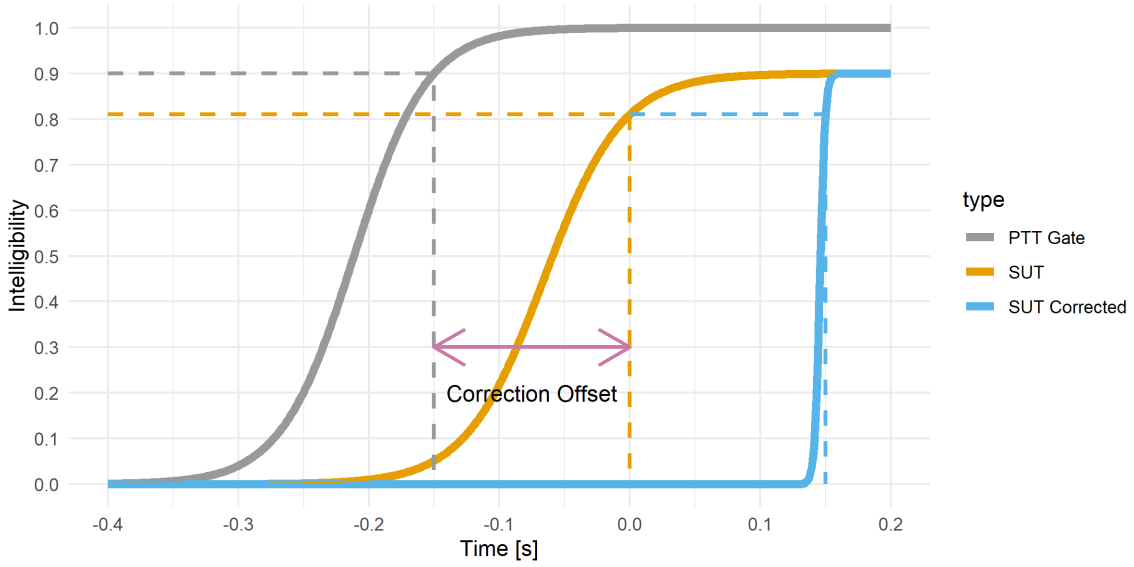

Fig. 2. Example intelligibility curves. The offset between the vertical dotted lines for PTT Gate and SUT represent the impact of the start-of-word correction on measured access delay. Note that the SUT Corrected curve has a much steeper slope than the SUT curve. This is an effect of the start-of-word correction removing word structure effects from access delay.

This motivates the need for a start-of-word correction using access delay results from the PTT gate. Adding a correction term from these results causes measurements on arbitrary SUTs to better reflect reality. In this example, 150 ms would be added to the access delay results so that $\tau_S(90\%) = 150$ ms. This accurately describes that when PTT is pressed right when the audio clip is sent to the transmit device, there is 150 ms of extraneous audio before the audio that drives a 90% intelligibility score is played. This extraneous audio is not part of the definition of access delay, so correcting the measurements to remove the effect from that additional audio better allows the access delay measurement system to reflect the definition of access delay.

4.2 Linear Combinations of Access Delays

Before getting into the derivation of the start-of-word correction, consider taking a linear combination of access delay functions. In particular, if we have N access delay functions, let $\tau_i(\alpha) = \lambda_i C + t_{0,i}$, where $C = \ln\left(\frac{1-\alpha}{\alpha}\right)$, with $i \in \{1,...,N\}$. If we now consider a linear combination of these access delays, $\tau(\alpha) = \sum_i^N a_i \cdot \tau_i(\alpha)$, it can be shown that

$$\tau(\alpha) = \left(\sum_{i=1}^{N} a_i \lambda_i\right) C + \left(\sum_{i=1}^{N} a_i t_{0,i}\right) \tag{6}$$

Thus, we can see that the linear combination of access delay functions is a function with a very similar form to an access delay function with new parameters. To be a meaningful access delay function, $\lambda < 0$ must hold, so one must always verify that $\left(\sum_{i=1}^{N} a_i \lambda_i\right) < 0$ when performing a linear combination of access delay functions.

4.3 Derivation

The start-of-word correction relies on using access delay from PTT gate measurements to correct access delays to account for critical intelligibility points within audio clips for each MRT keyword. In particular, the corrected access time for a given WUT in a SUT is defined as

$$\tau_{\mathrm{C}}(\alpha) = \tau_{\mathrm{S}}(\alpha) - \tau_{\mathrm{P}}(\alpha), \tag{7}$$

where $\tau_{\mathrm{S}}(\alpha)$ denotes the α level access delay for the SUT and $\tau_{\mathrm{P}}(\alpha)$ denotes the α level access delay for the PTT gate test. With Eq. (6), Eq. (7) can be rewritten as

$$\tau_{\mathrm{C}}(\alpha) = (\lambda_{\mathrm{S}} - \lambda_{\mathrm{P}}) \ln\left(\frac{1-\alpha}{\alpha}\right) + (t_{0,\mathrm{S}} - t_{0,\mathrm{P}}) \tag{8}$$

which will produce the corrected access times from the curve fit parameters of the SUT and PTT gate test. Note that as discussed in Sec. 3.2, $\lambda_{\mathrm{S}} < \lambda_{\mathrm{P}}$ is true for any nontrivial[2] SUT, so $(\lambda_{\mathrm{S}} - \lambda_{\mathrm{P}}) < 0$ will always be true, and the corrected access delay function is meaningful.

In general, access delay measurements use a variety of keywords and talker combinations, meaning multiple audio clips are used to characterize the access delay of a given SUT. The start-of-word correction needs to be performed on each WUT used in an access delay measurement for a SUT. The final access delay is then the average of all the corrected access delays. More explicitly, the access delay for a SUT when N words are used is

$$\tau_{\mathrm{sys}}(\alpha) = \frac{1}{N} \sum_{j=1}^{N} \tau_{\mathrm{C},j}(\alpha) \tag{9}$$

where $\tau_{\mathrm{C},j}(\alpha)$ is the corrected access time of the jth word.

[2]We consider any SUT that does not reasonably approximate the minimal access delay system of Sec. 3.1 to be nontrivial.

Using Eq. (6) and Eq. (8), Eq. (9) can be written as

$$\tau_{\text{sys}}(\alpha) = \frac{1}{N}\sum_{j=1}^{N}(\lambda_{j,\text{C}})\ln\left(\frac{1-\alpha}{\alpha}\right) + \frac{1}{N}\sum_{j=1}^{N}(t_{0,j,\text{C}}) \tag{10}$$

with $\lambda_{j,\text{c}} = \lambda_{j,\text{S}} - \lambda_{j,\text{P}}$ and $t_{0,\text{j,c}} = t_{0,j,\text{S}} - t_{0,j,\text{P}}$ being the corrected λ and t_0 of the jth word. Note that in Eq. (10), coefficients are just the averages of the corrected parameters for each WUT, so the system access delay is

$$\tau_{\text{sys}}(\alpha) = \bar{\lambda}\ln\left(\frac{1-\alpha}{\alpha}\right) + \bar{t}_0 \tag{11}$$

The above can be inverted to show that after correcting and averaging access delays for multiple words, a system still has a representative intelligibility curve such that

$$I_{\text{sys}}(t) = \frac{\bar{I}_0}{1+e^{(t-\bar{t}_0)/\bar{\lambda}}} \tag{12}$$

where $\bar{I}_0$ is the average of the asymptotic intelligibilities across words. This result is significant, as it allows us to interpret aggregated results from multiple WUTs on a single SUT the same way we do for any individual curve fit.

4.4 Interpretation of Aggregated Curve Parameters

The curve parameters of an intelligibility curve for a single word are all very interpretable, as has already been discussed. However, it is not obvious that when correcting multiple WUTs and averaging access delay results together, that we would ever get back to an interpretable intelligibility curve. Equations (11) and (12) show that we can define an intelligibility curve from the access delay function of averaged, corrected results. Here we will discuss how to interpret curve parameters that have been corrected and averaged.

For t_0, both the correction and averaging impact the final result the same way. Namely, changes to t_0 simply shift the final curve along the time axis, and the final t_0 value still describes the intelligibility transition midpoint, where $I_{\text{sys}}(t_0) = \bar{I}_0/2$.

When a given WUT is corrected, the corrected value λ_c removes word-based variability from the final curve. In particular, λ_S describes the access delay variability in the SUT and the steepness of the WUT's intelligibility transition. λ_P describes just the steepness of the WUT's intelligibility transition. This means the corrected value, λ_C, describes only the steepness of the intelligibility transition that is introduced by the SUT and is independent from word structure effects. When averaging corrected λ_C values, the result only describes the average transition time for the SUT. It is worth noting that individual corrected λ_C values will be different, even though we claim here that word-specific transition effects are eliminated. This is because a given SUT's performance is still WUT dependent; any given SUT may behave differently when it encounters different WUTs as stimuli.

The interpretation of the average of I_0 values is fairly natural, it is the average asymptotic intelligibility of the SUT across all the WUTs. This is analogous to how intelligibility for a SUT is normally defined.

Fig. 2 also plots the corrected intelligibility curve of the SUT. It can be seen that the $\tau(90\%)$ point on the corrected curve corresponds to the correction offset. It is also immediately clear that the intelligibility transition is much quicker for the corrected curve; in other words, λ_C is smaller in magnitude. This is due to the word-specific transition time being corrected with the PTT gate results. The transition time for the corrected curve is predominantly driven by the SUT. Finally, for this case of a single WUT, it is easy to see that I_0 remains unchanged from the original SUT curve to the corrected one.

4.5 Uncertainty Calculations

The uncertainty in access time measurements is derived from the formula estimating access time from the logistic fit given in Eq. (2). An access delay estimate $\hat{t}$ for a given value of α is defined as

$$\hat{t} = \ln\left(\frac{1-\alpha}{\alpha}\right)\hat{\lambda} + \hat{t}_0$$

Letting $C = \ln\left(\frac{1-\alpha}{\alpha}\right)$ and taking the variance from both sides of the equation yields:

$$\mathrm{Var}(\hat{t}) = C^2\mathrm{Var}(\hat{\lambda}) + \mathrm{Var}(\hat{t}_0) + 2C\cdot\mathrm{Cov}(\hat{\lambda},\hat{t}_0) \tag{13}$$

Since C is a function of α, the variance of the access time measurement is determined from α, as the remaining variance and covariance quantities are determined from the fitting algorithm. The uncertainty of $\hat{t}$ is determined in the usual way as $u(\hat{t}) = \sqrt{\mathrm{Var}(\hat{t})}$. The fitting algorithm used is *minpack.lm* package in the ***R*** programming language [8].

In context of propagating uncertainty between words, the parameter estimates for individual WUTs are independent, implying that their covariance terms are zero. This means that we can use Eq. (10) and take the variance on both sides of the equation. Again, letting $C = \ln\left(\frac{1-\alpha}{\alpha}\right)$ with the rules of variance and covariance addition for the access delay estimate of a SUT, $\hat{t}_{\mathrm{sys}}$, we have

$$\mathrm{Var}(\hat{t}_{\mathrm{sys}}) = \frac{1}{N^2}\left(C^2\sum_{j=1}^{N}\mathrm{Var}(\hat{\lambda}_{j,\mathrm{C}}) + \sum_{j=1}^{N}\mathrm{Var}(\hat{t}_{0,j,\mathrm{C}}) + 2C\sum_{j=1}^{N}\mathrm{Cov}(\hat{\lambda}_{j,\mathrm{C}},\hat{t}_{0,j,\mathrm{C}})\right) \tag{14}$$

where

$$\begin{aligned}
\mathrm{Var}(\hat{\lambda}_{j,C}) &= \mathrm{Var}(\hat{\lambda}_{j,\mathrm{S}}) + \mathrm{Var}(\hat{\lambda}_{j,\mathrm{P}}) \\
\mathrm{Var}(\hat{t}_{0,j,C}) &= \mathrm{Var}(\hat{t}_{0,j,\mathrm{S}}) + \mathrm{Var}(\hat{t}_{0,j,\mathrm{P}}) \\
\mathrm{Cov}(\hat{\lambda}_{j,\mathrm{C}},\hat{t}_{0,j,\mathrm{C}}) &= \mathrm{Cov}(\hat{\lambda}_{j,\mathrm{S}},\hat{t}_{0,j,\mathrm{S}}) + \mathrm{Cov}(\hat{\lambda}_{j,\mathrm{P}},\hat{t}_{0,j,\mathrm{P}})
\end{aligned} \tag{15}$$

This gives the equation for calculating uncertainty in the system access delay $\tau_{\mathrm{sys}}(\alpha)$.

4.6 Implication of Corrections in Measurements

In the original measurement system, a single curve was fit across data from a variety of WUTs. In the new curve fitting regime, a curve is fit to the intelligibility data from each individual WUT. This requires that each WUT is well-behaved from an intelligibility estimation perspective. In practice, it has been observed that certain MRT keywords are more prone to erroneous intelligibility estimations when part of the beginning of the keyword is muted, as happens in access delay measurements. Thus keyword selection for access delay measurements must be well considered, and some searching needs to be done to ensure that keywords behave well in both PTT gate tests and a variety of SUTs. This is discussed further in Sec. 5.1.

5. Example of Results

In this section, we present some example measurement results to demonstrate the impact the start-of-word correction has on access delay. In Ref. [1], a variety of land mobile radio (LMR) technologies and configurations were tested. Here we performed new measurements on a similar selection of technologies, including an in-development Long Term Evolution (LTE) system to emphasize that this measurement system works on any PTT communications system. In order to demonstrate the impact of the start-of-word correction, results will also be shown with the original curve fitting procedure from Ref. [1]. Please note that all curve parameters and measurement results are presented with 95% confidence intervals included in parentheses. All confidence intervals use a coverage factor of $k = 1.96$ to generate an expanded uncertainty from the standard uncertainty.

5.1 Keyword Selection

With the start-of-word correction, it is imperative that an intelligibility curve can be fit to each individual keyword. For certain technologies, some keywords exhibit highly variable intelligibilities for both the measurements of P_1 when the time between pressing PTT and speaking the keyword was large and of P_2. It was observed that in these instances, the intelligibility curve fits for individual words were often not consistent nor accurate. This behavior was masked by the previous curve fitting methodology, where a single curve was fit to the data from all WUTs at once. This motivated us to revisit keyword selection for access delay measurements.

The keywords originally selected for access delay measurements in Ref. [1], were chosen by identifying the keywords that minimized the root mean square error (RMSE) between the average ABC-MRT score of the four selected keywords and the average ABC-MRT score when 1200 MRT keywords are used. The RMSE operation was performed across the 139 radio and codec conditions described in Ref. [9]. The RMSE for the originally selected words corresponds to the smallest RMSE possible under the conditions that one word is selected from each of four speakers, and words are only drawn from batches with differing leading consonants.

This procedure selects keywords that provided more accurate estimates of a technology's asymptotic intelligibility, but makes no guarantees on yielding an obvious intelligibility transition when varying PTT time. The intelligibility transition of a keyword from 0 to its asymptotic intelligibility is what we aim to measure when calculating access delay, not the asymptotic intelligibility itself. To this effect we realized this transition would be easiest to consistently measure for keywords that have high intelligibility with low variance when sent through different codecs without any muting of the keyword.

For new keyword selection, we again focused on keywords from batches where the leading consonant varies. We performed analysis with the software implementations of the following codecs: analog FM, P25 FR, P25 HR, AMRNB7 (12.20 kbps), AMRWB2 (12.65 kbps). For each keyword and codec combination, 120 trials were performed. Audio for each trial had a unique noise signal at constant signal-to-noise ratio (SNR), and had a random amount of silence prior to speech; this randomized the keyword's location within frame alignment for the digital codecs. For each talker, we identified the keywords with lowest variance and highest intelligibility across all trials for all codecs. Each talker had multiple keywords that achieved intelligibility of 1 with variance of 0.

After generating lists of candidate words for each talker, we ran access delay tests with the PTT gate as a final verification step. Some candidate keywords had erroneous intelligibility estimates in the access delay measurement and were eliminated. We further constrained our selections so that the keyword for each talker was unique. This resulted in us selecting the keywords in Table 1. Four keywords, one per talker, are selected to have a variety of different talkers while also keeping test time to a reasonable limit. For convenience from here on, we refer to these WUTs only by the talker and the actual keyword, e.g. F1 bed.

Table 1. Selected keywords for each talker. Batch refers to the MRT list number, and word number designates the word in the list [3].

Talker	Batch	Word Number	Word
F1	9	1	bed
F3	31	2	law
M3	38	1	hang
M4	14	1	not

5.2 Correction Curves

Here we present intelligibility curves for the access delay start-of-word correction for the selected keywords. These curves come from access delay measurements using the PTT gate. The curve parameters can be seen in Table 2 and the intelligibility curves can be seen in Fig. 3. Note that these curves are fully defined by the parameters in Table 2, and can be used to correct any access delay measurement that uses these keywords on any SUT.

Table 2. Access delay of PTT Gate, 95% confidence intervals presented in parentheses. Note that I_0 was measured to be identically 1, so the confidence interval is trivial.

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	1, (1, 1)	-0.04508, (-0.04514, -0.04502)	-0.0028938, (-0.0029276, -0.0028603)
F3 law	1, (1, 1)	-0.163, (-0.164, -0.163)	-0.0244, (-0.0251, -0.0237)
M3 hang	1, (1, 1)	-0.07553, (-0.07581, -0.07525)	-0.01003, (-0.01025, -0.009812)
M4 not	1, (1, 1)	-0.081551, (-0.081589, -0.081513)	-0.0036713, (-0.0037115, -0.0036309)

More time resolution is required to achieve accurate and repeatable measurements when fitting intelligibility curves to individual keywords using the PTT gate. This is due to the quick turn-on time and lack of variability in the PTT gate. A time step size of 10 ms rather than the typical 20 ms is used in all PTT gate access delay measurements. Further, measurements on the PTT gate require the use of firmware version 1.2.2 or higher for the radio interface [10]. That version improves clock accuracy, which makes PTT timing in access delay measurements more accurate; this in turn significantly improves the repeatability of PTT gate access delay results.

5.3 Impact on Access Delay

Example access delay measurements were performed on the following technologies: analog FM direct mode, an in-development LTE system, Project 25 (P25) direct mode, P25 Phase 1 trunked, and P25 Phase 2 trunked. The results from these example measurements can be seen in Table 3. Access delay results that are measured and corrected using the procedure described in Sec. 4.3, are in this table, labeled as Corrected Access Delay. The table also shows the access delay results computed without the start-of-word correction and using the old curve fitting procedure of Ref. [1], where a single curve was fit to all the data across all WUTs at once. In the table, these results are labelled as Legacy Access Delay. The difference between these two methodologies is also shown. As was expected, the start-of-word correction increases access delay measurements; access delays increase by between 15 ms to 44 ms in the example measurements performed here.

Figure 4 shows comparison plots of corrected versus legacy access delay results. For every technology, the corrected access delay curve is much flatter than the legacy curve. This is primarily because the correction attempts to remove word-based effects from the measurement. In other words, the intelligibility transition is being driven completely by the SUT, so variation from the WUTs is removed, and the transition from low to high intelligibility is quicker. It is also easy to see that for each technology, the corrected curve has higher access delay across all α values until α gets close to 1. Again this is because the

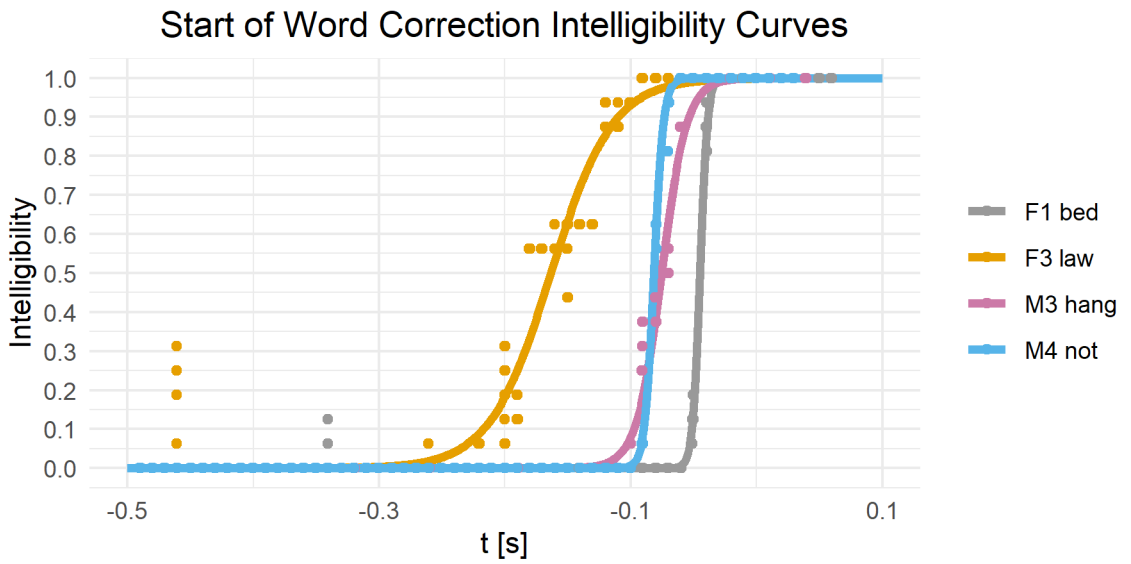

Fig. 3. Start-of-word correction intelligibility curves. The raw data used to calculate these curves are also plotted.

start-of-word correction appropriately accounts for nonessential speech within the audio clips containing the MRT keywords.

Finally, Fig. 5 shows access delay as a function of intelligibility for each technology. Note that when compared to Fig. 16b in Ref. [1], the asymptotic intelligibilities for all technologies are higher. This does not represent an improvement in the intelligibility of these systems, and is instead driven by the new keyword selection procedure described in Sec. 5.1. It is worth noting that the start-of-word correction does not affect the ordering of the magnitude of access delay for each technology reported previously; when using the same WUTs, the same correction is added for every technology at a given α level.

6. Conclusion

A start-of-word correction for access delay measurements was implemented. The example measurements demonstrate that the measurement system works on a variety of technologies and that the start-of-word correction improves access delay estimates by appropriately representing audio immaterial to intelligibility in the final output. The example measurements demonstrate the start-of-word correction and new curve fitting methodology increase access delay estimates on the order of 10 ms to 50 ms. The start-of-word correction successfully refocuses access delay measurements to represent SUT effects by removing emphasis from WUT effects. Adding the start-of-word correction is an important improvement to the access delay measurement system that ensures measurement results more accurately reflect access delay from a user experience perspective.

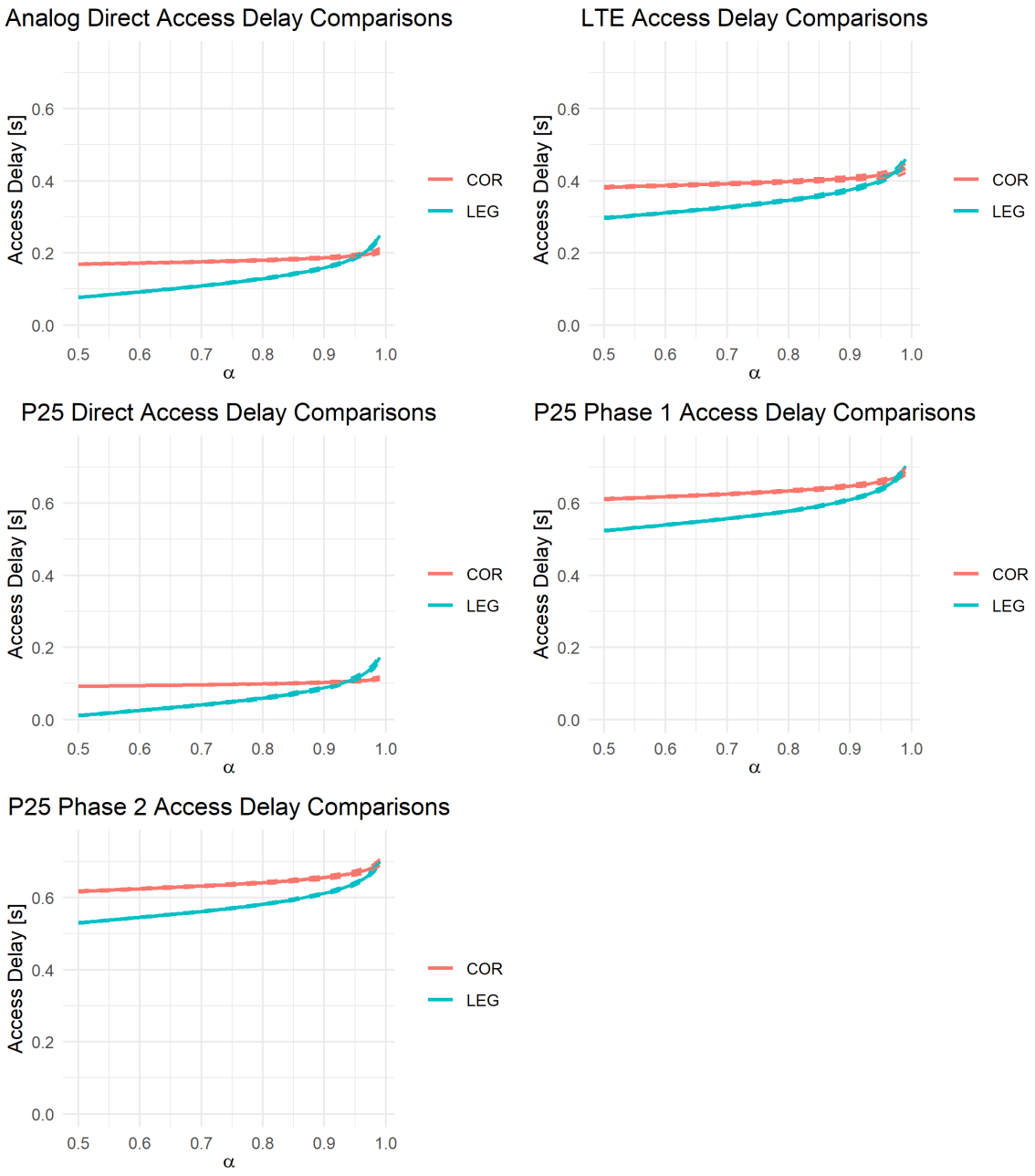

Fig. 4. Access delay comparisons. COR denotes access delay measurements using the start-of-word correction, where LEG denotes legacy measurements relying on a single curve fit.

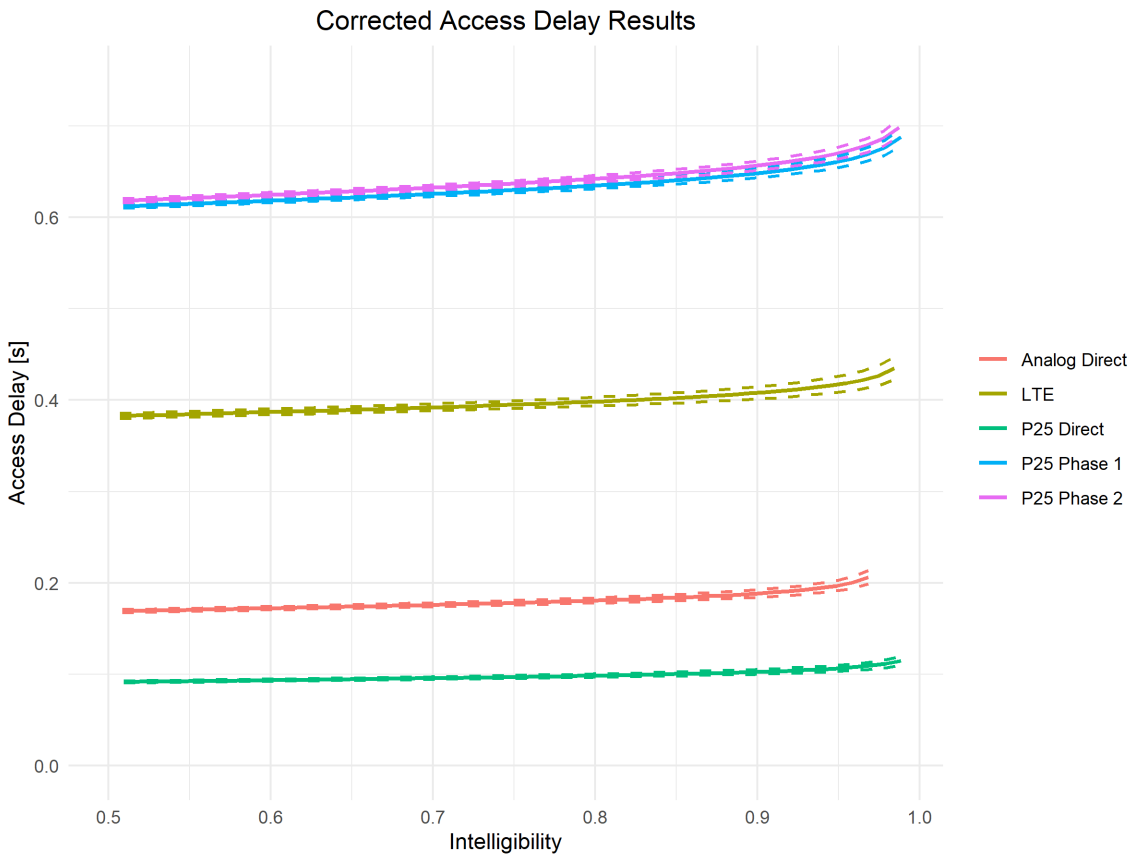

Fig. 5. Access delay results with start-of-word correction. Dotted lines represent 95% confidence intervals.

Table 3. Start-of-word correction access delay impact. Corrected access delay refers to access delay values using the start-of-word correction. Legacy access delay refers to access delay using the curve fitting procedure originally introduced in Ref. [1]. 95% confidence intervals are included in parentheses. All access delays are calculated with $\alpha = 0.9$.

Technology	Corrected Access Delay [ms]	Legacy Access Delay [ms]	Difference [ms]
Analog Direct	186.5, (184.6, 188.4)	159.3, (157.5, 161.1)	27.2, (24.5, 29.8)
LTE	407.3, (404.0, 410.6)	375.4, (372.7, 378.1)	31.9, (27.6, 36.2)
P25 Direct	102.7, (101.4, 104.1)	87.92, (85.61, 90.22)	14.8, (12.2, 17.5)
P25 Trunked Phase 1	647.9, (645.4, 650.4)	610.2, (608.3, 612.2)	37.7, (34.5, 40.9)
P25 Trunked Phase 2	656.1, (653.5, 658.6)	612.3, (610.5, 614.0)	43.8, (40.7, 47.0)

References

[1] Pieper J, Frey J, Greene C, Soetan Z, Thompson T, Voran S, Bradshaw D (2019) Mission Critical Voice QoE Access Time Measurement Methods (NIST), IR-8275. doi: 10.6028/NIST.IR.8275

[2] Greene C, Frey J, Soetan Z, Pieper J, Thompson T (2020) Mission Critical Voice Quality of Experience Access Time Measurement Method Addendum (NIST), IR-8328. doi: 10.6028/NIST.IR.8328

[3] ANSI/ASA (2009) ANSI/ASA S3.2-2009 Method for Measuring the Intelligibility of Speech over Communication Systems.

[4] Voran SD (2017) A multiple bandwidth objective speech intelligibility estimator based on articulation index band correlations and attention. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, , pp 5100–5104. doi: 10.1109/ICASSP.2017.7953128

[5] (2016) Modified rhyme test audio library, NIST. URL https://www.nist.gov/ctl/pscr/modified-rhyme-test-audio-library.

[6] TIA (2016) Digital C4FM/CQPSK Transceiver Measurement Methods (Telecommunications Industry Association (TIA)), Standard 102.CAAA-E.

[7] 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.

[8] Elzhov TV, Mullen KM, Spiess AN, Bolker B (2016) *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*, . R package version 1.2-1 URL https://CRAN.R-project.org/package=minpack.lm.

[9] Voran SD (2013) Using articulation index band correlations to objectively estimate speech intelligibility consistent with the modified rhyme test. *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE), , pp 1–4.

[10] Frey J (2019–2021) Radio Interface Firmware. URL https://github.com/usnistgov/MCV-QoE-firmware.

Appendix

A. Example Measurement Results

Curve parameters for all the example measurements are included here. Here aggregated means that effects from all four WUTs are included. Note that 95% confidence intervals are presented in parentheses.

Table 4. Corrected curve parameters for example analog direct measurements.

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	0.9999, (0.9998, 1)	0.153, (0.15, 0.155)	-0.0248, (-0.0273, -0.0223)
F3 law	0.975, (0.974, 0.976)	0.173, (0.172, 0.175)	-0.00535, (-0.00688, -0.00382)
M3 hang	0.9376, (0.9374, 0.9378)	0.216, (0.215, 0.216)	-0.0017, (-0.0023, -0.00109)
M4 not	1, (1, 1)	0.132, (0.131, 0.133)	-0.000877, (-0.0016, -0.000152)
*aggr**	0.9782, (0.9779, 0.9785)	0.169, (0.167, 0.17)	-0.00818, (-0.00973, -0.00663)

* Indicates aggregated technology curve

Table 5. Corrected curve parameters for example P25 direct measurements.

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	0.995, (0.994, 0.996)	0.105, (0.105, 0.106)	-0.00679, (-0.00735, -0.00623)
F3 law	1, (1, 1)	0.0615, (0.0595, 0.0636)	-0.0122, (-0.014, -0.0104)
M3 hang	0.999, (0.999, 1)	0.113, (0.112, 0.114)	-0.000365, (-0.000908, 0.000178)
M4 not	1, (1, 1)	0.086, (0.08568, 0.08632)	-0.00111, (-0.0019, -0.000318)
*aggr**	0.9987, (0.9984, 0.999)	0.0915, (0.0904, 0.0926)	-0.00511, (-0.00617, -0.00404)

* Indicates aggregated technology curve

Table 6. Corrected curve parameters for example P25 Phase 1 trunked measurements.

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	0.996, (0.995, 0.997)	0.62, (0.617, 0.623)	-0.0229, (-0.0253, -0.0205)
F3 law	1, (1, 1)	0.602, (0.6, 0.604)	-0.0079, (-0.00987, -0.00592)
M3 hang	0.9998, (0.9995, 1)	0.613, (0.611, 0.614)	-0.0162, (-0.0178, -0.0145)
M4 not	0.998, (0.996, 1)	0.611, (0.609, 0.613)	-0.0195, (-0.0214, -0.0177)
*aggr**	0.9986, (0.998, 0.999)	0.611, (0.609, 0.614)	-0.0166, (-0.0186, -0.0146)

* Indicates aggregated technology curve

Table 7. Corrected curve parameters for example P25 Phase 2 trunked measurements

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	0.992, (0.989, 0.994)	0.614, (0.612, 0.617)	-0.0209, (-0.0233, -0.0186)
F3 law	1, (0.9999, 1)	0.614, (0.612, 0.616)	-0.00715, (-0.00868, -0.00561)
M3 hang	0.99998, (0.99993, 1)	0.626, (0.623, 0.628)	-0.0187, (-0.0206, -0.0167)
M4 not	0.998, (0.996, 1)	0.615, (0.613, 0.618)	-0.0236, (-0.0259, -0.0214)
*aggr**	0.997, (0.997, 0.998)	0.617, (0.615, 0.62)	-0.0176, (-0.0196, -0.0156)

* Indicates aggregated technology curve

.

Table 8. Corrected curve parameters for example LTE measurements.

WUT	I_0	t_0 [s]	λ [s^{-1}]
F1 bed	0.988, (0.983, 0.993)	0.383, (0.378, 0.387)	-0.0101, (-0.0137, -0.00647)
F3 law	0.9998, (0.9996, 1)	0.372, (0.369, 0.375)	-0.00971, (-0.0122, -0.00725)
M3 hang	0.999, (0.998, 1)	0.398, (0.395, 0.401)	-0.0124, (-0.0148, -0.00997)
M4 not	0.991, (0.987, 0.995)	0.377, (0.375, 0.379)	-0.0135, (-0.0152, -0.0117)
*aggr**	0.995, (0.993, 0.996)	0.382, (0.379, 0.385)	-0.0114, (-0.0141, -0.00877)

* Indicates aggregated technology curve