PSCR 2021

THE DIGITAL EXPERIENCE

#PSCR2021 • PSCR.GOV

NIST

PSCR

Introducing a Start of Word Correction for Access Delay Measurements

Jaden Pieper PSCR MCV

William Magrogan PSCR MCV

#PSCR2021

NIST

PSCR

DISCLAIMER

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept 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 entities, materials, or equipment are necessarily the best available for the purpose.

*** Please note, unless mentioned in reference to a NIST Publication, all information and data presented is preliminary/in-progress and subject to change**

Acronym Index

- Mission Critical Voice (MCV)
- System Under Test (SUT)
- Word Under Test (WUT)
- Push To Talk (PTT)
- Key Performance Indicator (KPI)
- Quality of Experience (QoE)
- Modified Rhyme Test (MRT)
- Articulation Band Correlated Modified Rhyme Test (ABC-MRT)

Overview

- Quality of Experience Based Measurements
- Access Time Measurement System
- Intelligibility
- Access Delay Model
- Start of Word Correction
- Interpretation
- Effect

Quality of Experience Based Measurements

- QoE KPIs for Mission Critical Voice (MCV)
  - Mouth-to-Ear Latency
  - Access Time
  - Voice Quality/Intelligibility
  - Probability of Successful Delivery

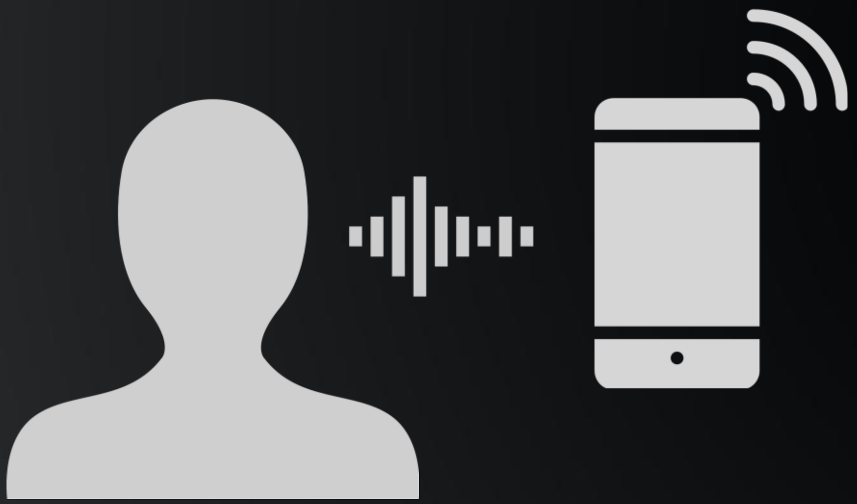
The User Experience: PTT Communications

- Press PTT and speak into a device
- Listening to speech output from a device
- It is all about <u>speech</u>

- Goal - Create measurement systems that are:
  - Based upon the user experience -- speech
  - Comparable and fair across technologies
- This is not:
  - Analyzing internal system design/construction

Technology Agnostic Measurements

Communications System



KPIs:

- Mouth-to-ear Latency
- Access Time
- Intelligibility
- Probability of Successful Delivery

User Driven Access Definition

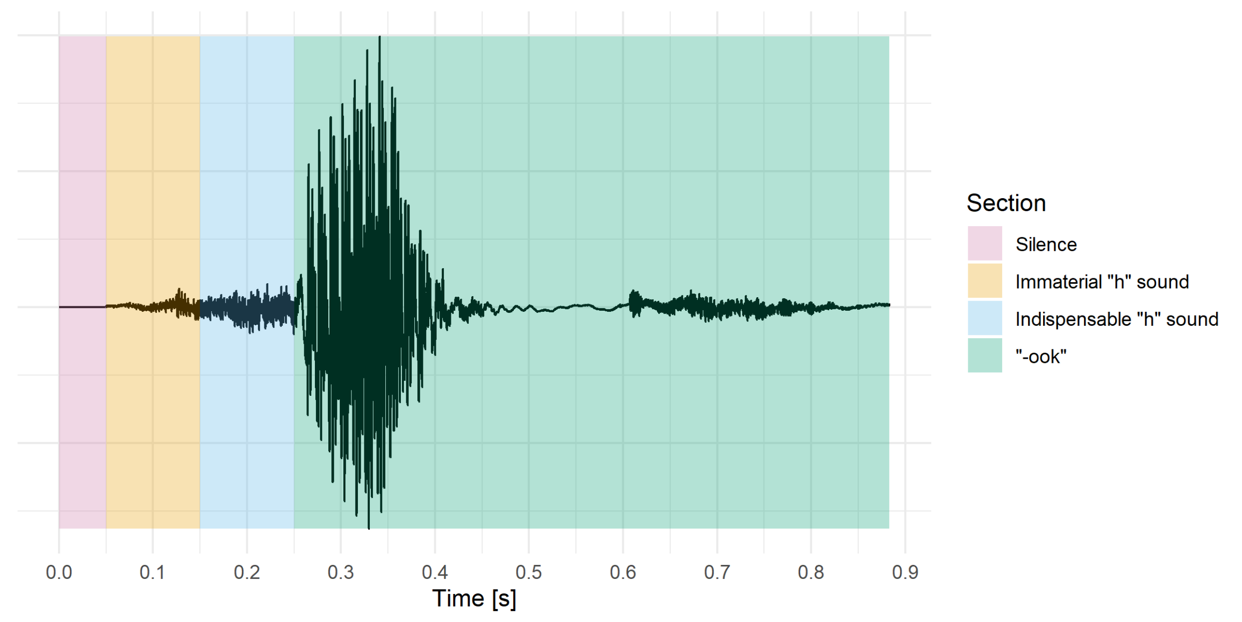
- End-to-end Access Time
  - *The total amount of time from when a transmitting user first presses PTT until a receiving user hears intelligible audio*
- Two components:
  - Mouth-to-Ear Latency
    - *The time between speech being input into one device and its output through another*
  - Access Delay
    - *The minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the start of the message is not lost*

Formalizing Access Delay

- Access Delay
  - All about if a message is lost or not
  - Intelligibility is the key to the measurement
- Formal definition
  - *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* $\alpha \cdot I_0$
  - $0 < \alpha < 1$, defines acceptable intelligibility level
  - I_0 is the baseline intelligibility of that word through the communications system
    - No system is perfectly intelligible
    - Some level of degradation almost always present

Start of Word Correction - Intuition

- Think about the word “hook”
  - Hear and choose from list of
    - took, cook, look, hook, shook, book
  - Long “h” sound, takes up 200 ms
  - Need 100 ms to understand “hook” from “took”, “look”, etc…
  - Audio clip: 50 ms of silence at beginning
- Problem
  - Currently this extra 150 ms gets misrepresented in access delay measurements



Intelligibility

- Modified Rhyme Test (MRT)
  - Batches of six words
    - *went, sent, bent, dent, tent, rent*
    - Words: consonant-vowel-consonant
    - Each batch: Either leading or trailing consonant varies
  - MRT Trial
    - Carrier phrase + word
    - e.g. “Please select the word *went*”
    - Success (identified) or Failure (mis-identified)
- ABC-MRT16
  - Objective algorithm to estimate MRT scores
  - Get results “on demand”
- For more information see:
  - Designing Remote Listening Experiments for the Partially Muted Word Impairment

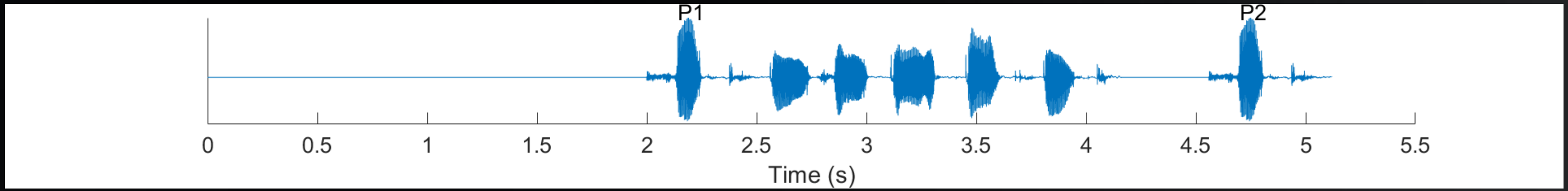
Measurement Concept

- Access Delay definition
  - *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* $\alpha \cdot I_0$
- Repeatedly send pre-defined audio clips through communications system
- Vary where in the clip PTT is triggered
- Measure relationship between PTT time and intelligibility of the first word in the clip
  - No more carrier phrase
  - “Please select the word west” → “west”
- Time between PTT and start of audio clip
  - Not necessarily the same as start of keyword
  - Need to account for this

Measurement Concept

Audio Clips

- Select single word from ABC-MRT16 database[1]
  - Use only words from batches where leading consonant varies
    - E.g. *went, sent, bent, dent, tent, rent*
    - Places majority of intelligibility emphasis on beginning of word
  - Remove carrier phrase
    - Cut out keyword conservatively
    - Extra audio at beginning of clip
- Structure:
  - T seconds of silence
  - Play word, P_1
  - T seconds of speech
  - Play word again, P_2

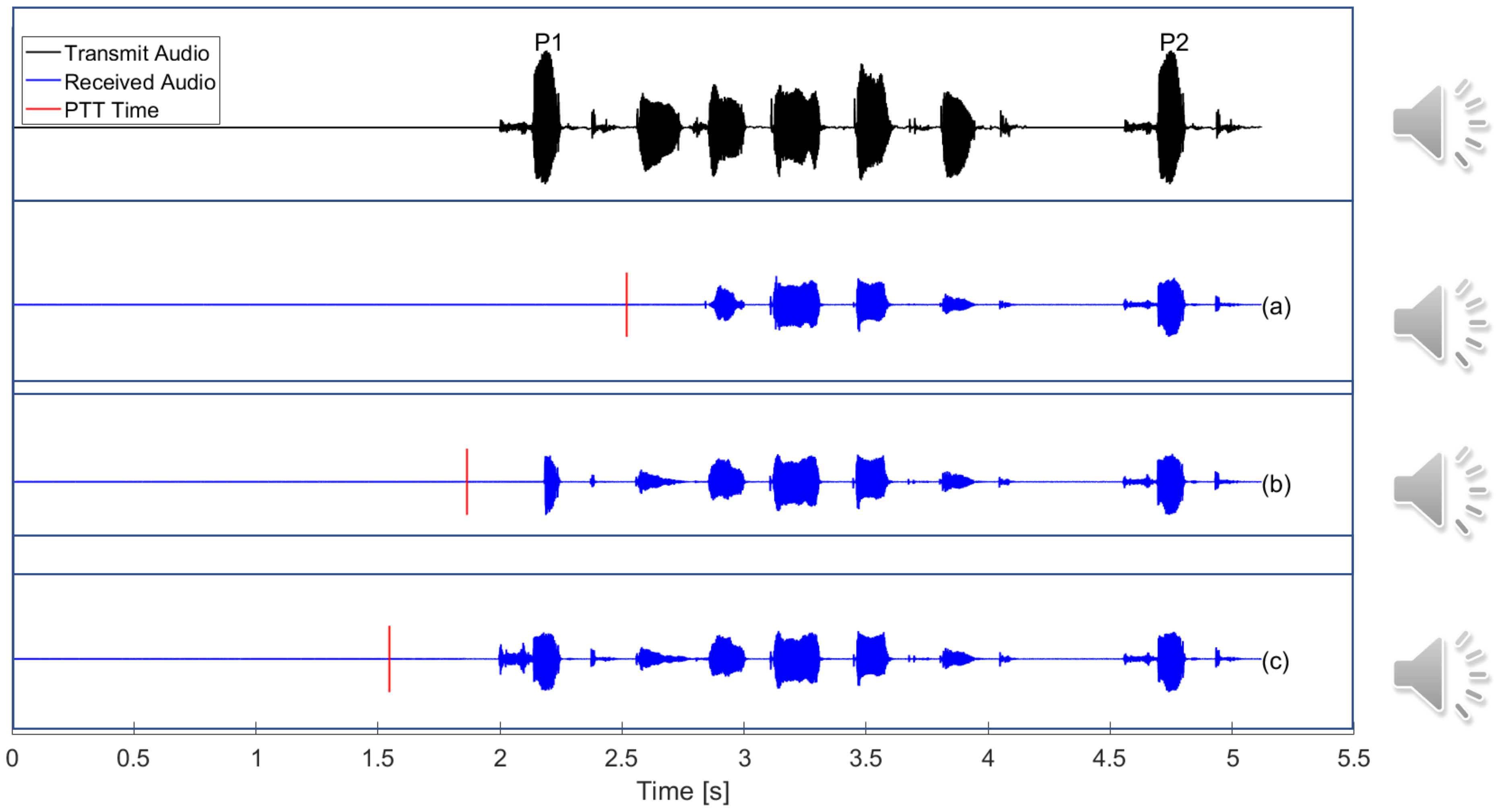


- T chosen so that system access time is less than T seconds
- Intelligibility of P_2 describes the asymptotic intelligibility, I_0
- Intelligibility of P_1 relates PTT time with intelligibility

1: 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

Measurement Concept

Intelligibility Examples: *hook*

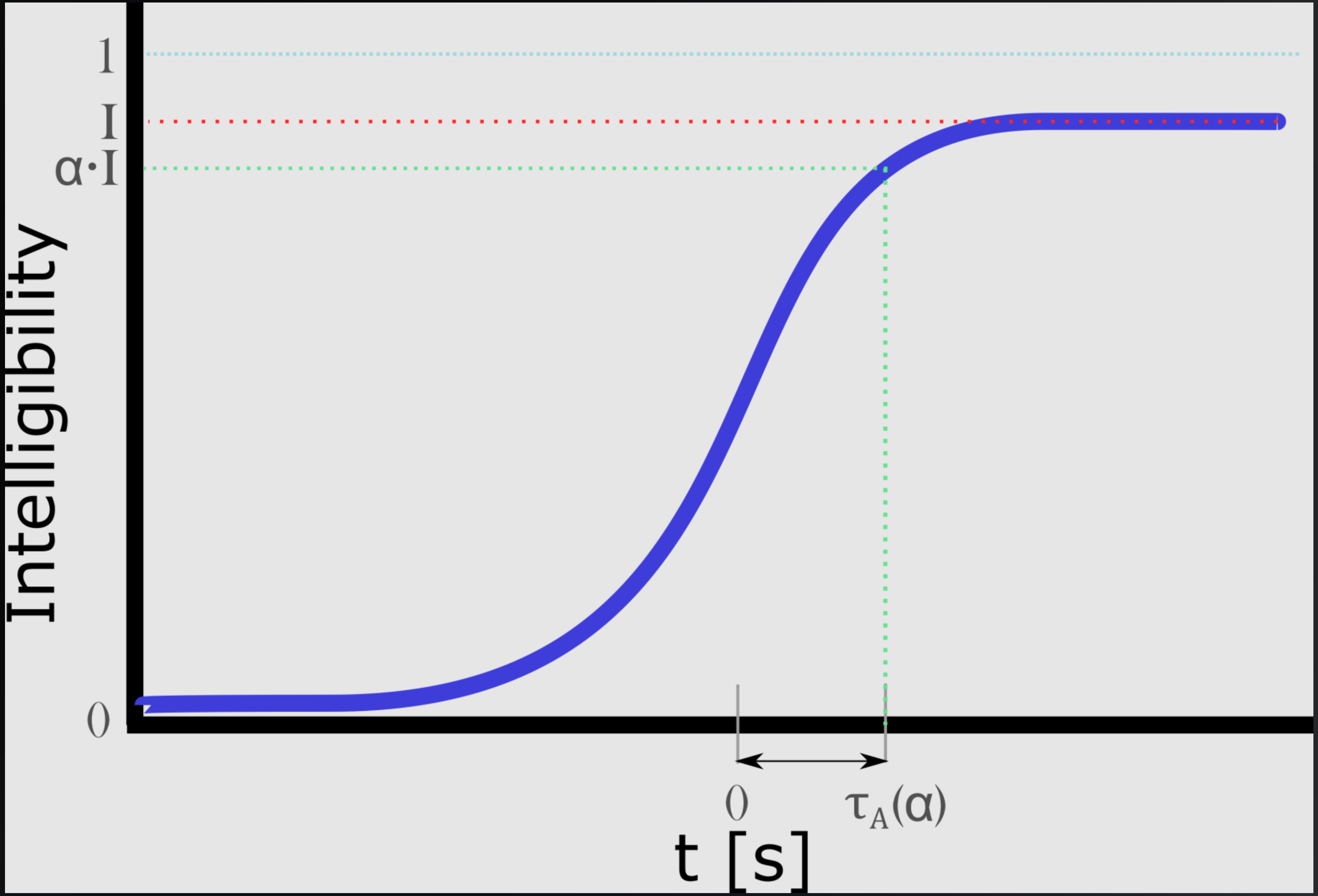


Measurement Concept

- Fit a curve to data
- Logistic curve has properties we want

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

- $\lambda < 0$, steepness of intelligibility transition
- $t_0 \in \mathbb{R}$, 50% intelligibility point



Measurement Concept

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

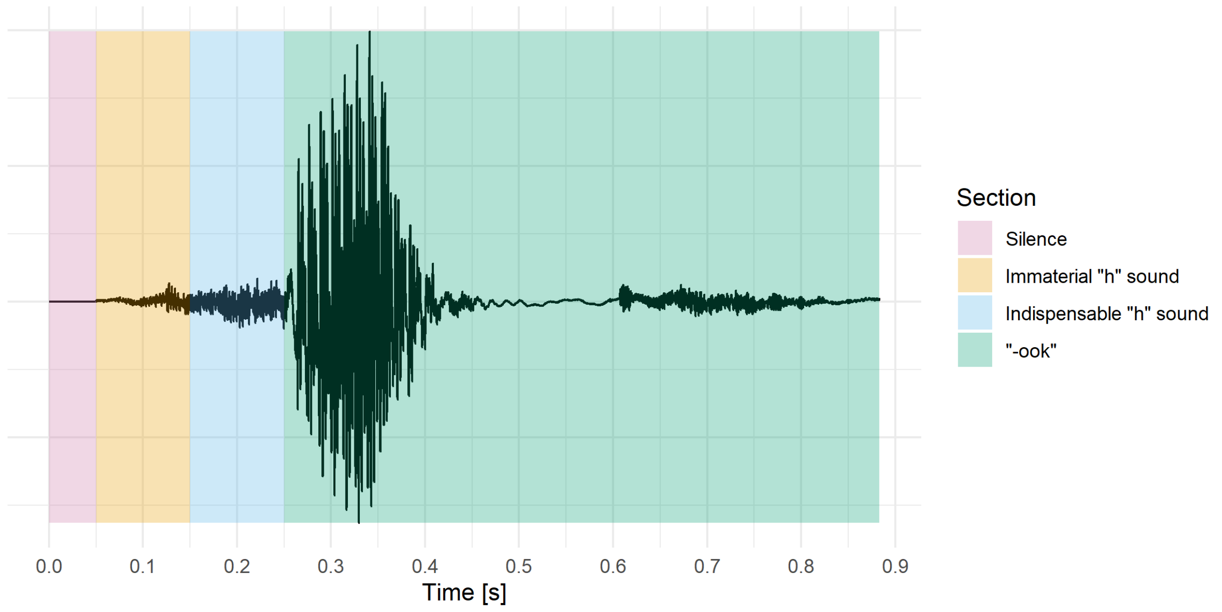
Given $0 < \alpha < 1$, an intelligibility of $\alpha \cdot I_0$ can be achieved with $t = I^{-1}(\alpha \cdot I_0)$

Access Delay defined as:

$$\tau_A(\alpha) = \lambda \cdot \ln\left(\frac{1-\alpha}{\alpha}\right) + t_0$$

Start of Word Correction - Question

- How do we characterize the immaterial from the indispensable?
- Example: “hook”
  - How much leading audio contributes to intelligibility?
  - Cutoff point changes for each word
  - Independent of SUT
  - Approximate an *ideal* SUT: PTT gate only



MRT List: took, cook, look, hook, shook, book

Minimal Access Delay System

- Consider a system, s^*, that minimizes access delay for a given WUT
  - Let S be the set of all SUTs that perform no audio buffering prior to PTT
  - Define $s^* \in S$ such that for any $\alpha \in \left[\frac{1}{2}, 1\right)$, then:
$$\tau_{s^*}(\alpha) \leq \tau_s(\alpha), \forall s \in S,$$
- Theorem: Access delay parameters λ_{s^*}, t_{0,s^*} provide maximum and minimum values, respectively, for the WUT through any SUT, $s \in S$
  - $\lambda_{s^*} \geq \lambda_s$
  - $t_{0,s^*} \leq t_{0,s}$

Minimal Access Delay System

- Minimal Access Delay
  - Minimizes transition time, the magnitude of λ, and offset, t_0
  - Helps discern information about the WUT
  - What is the best practical approximation?

PTT Gate

- What is it?
  - Input → Digital Switch → Output
  - ~ 90µs transition time
- Approximately minimal system
  - PTT gate and nothing else
  - No system effects (squelch, channel access, etc)
  - Access delay always negative
  - Provides baseline access delay per WUT
- Prohibits access delay hiding in system delay
- Correcting data
  - How to use for correcting data?

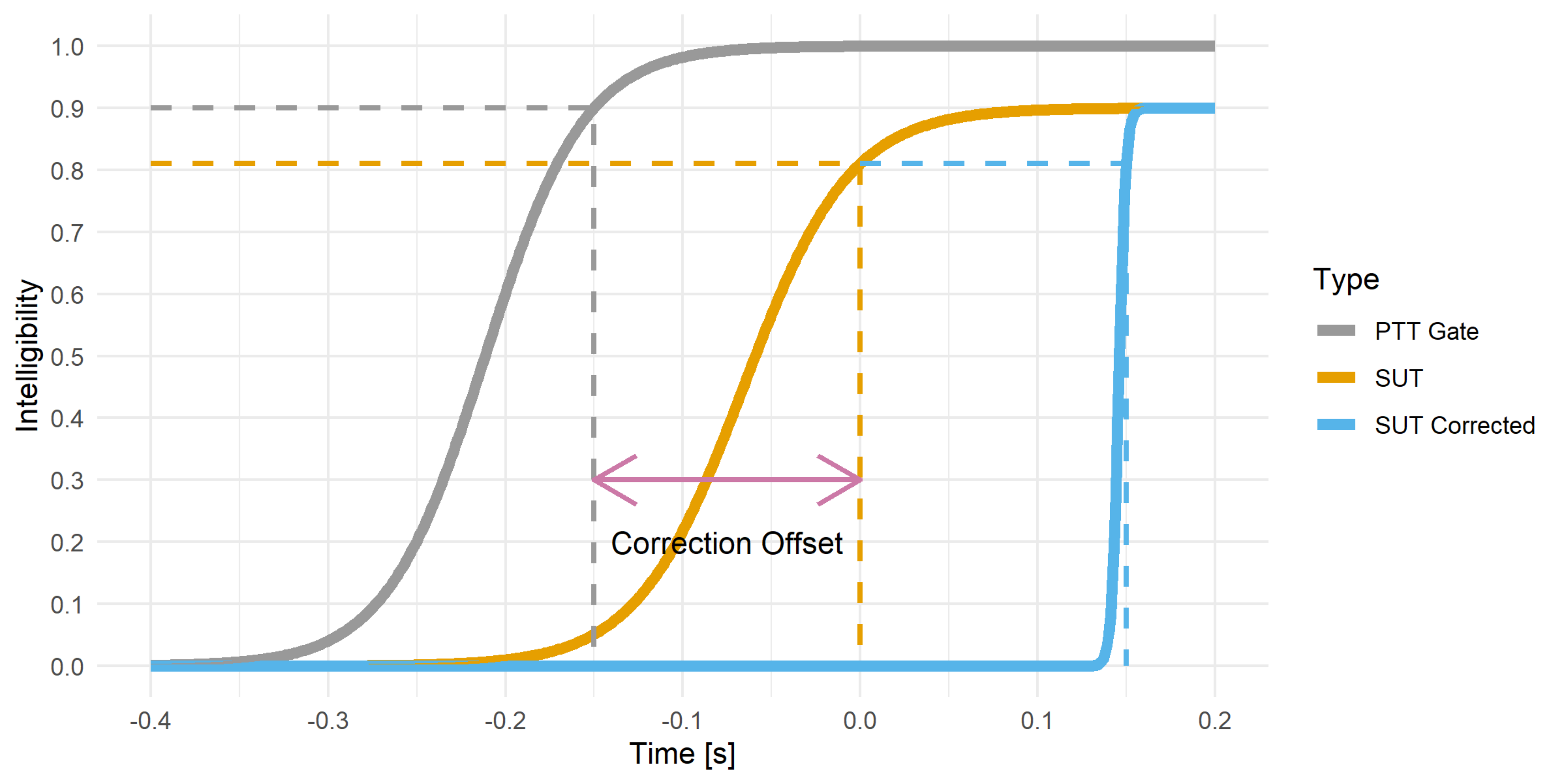
Linear Combinations of Access Delay

- Access delay functions have nice properties
  - Adding Access Delay → Adding Parameters
- Hope you like math
  - Given some fixed α
  - Invert logistic curve for access delay function
  - $\tau = \ln\left(\frac{1-\alpha}{\alpha}\right)\lambda + t_0$ for one word
  - Combine as usual
  - $a\,\tau_1 + b\,\tau_2 = \ln\left(\frac{1-\alpha}{\alpha}\right)(a\lambda_1 + b\lambda_2) + (at_{0,1} + bt_{0,2})$
  - $a\lambda_1 + b\lambda_2 < 0$ to be valid

Corrected Access Delay

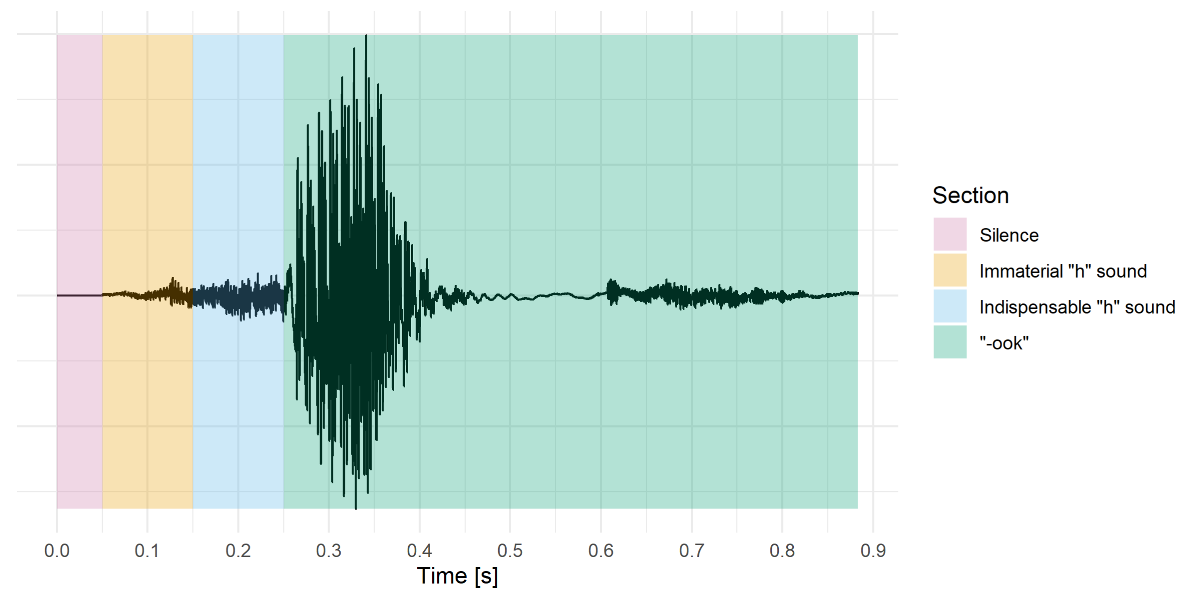
- Remove effects of immaterial audio
  - PTT gate tests characterize WUT related effects
  - Subtract off access delay from PTT gate for given WUT
    - $\tau_{SUT} - \tau_{PTT} = \ln\left(\frac{1-\alpha}{\alpha}\right)(\lambda_{SUT} - \lambda_{PTT}) + \left(t_{0,SUT} - t_{0PTT}\right)$
  - Corrected access delay characterizes SUT only effects
- Average over different WUTs
  - $\tau_c = \frac{1}{N}\sum(\tau_{SUT} - \tau_{PTT}) = \ln\left(\frac{1-\alpha}{\alpha}\right)\frac{1}{N}\sum(\lambda_{SUT} - \lambda_{PTT}) + \frac{1}{N}\sum\left(t_{0,SUT} - t_{0,PTT}\right)$
  - $\lambda_{PTT} > \lambda_{SUT}$ implies valid parameters
  - Multiple WUT used to sample how SUT responds to different words/talkers
- Corrected access delay times also follow logistic curve

Corrected Access Delay - Example



Start of Word Correction - Example

- Fix $\alpha = 0.9$
- Think about the word "hook"
- $\tau_{PTT}(\alpha) = -150$ ms
- $\tau_{SUT}(\alpha) = 0$ ms
  - Before correction: no access delay
  - Know that the first 150 ms of the audio clip "worthless"
- $\tau(\alpha) = \tau_{SUT}(\alpha) - \tau_{PTT}(\alpha) = 0 - -150 = 150$ ms
  - Captures the "truth" better
  - System needs 150 ms of setup time to send "hook" through intelligibly



Interpretation

- For single word
  - t_0 will shift intelligibility midpoint and become larger
  - λ will remain negative and become smaller in size (less negative)
    - This is true because the PTT gate approximates minimal access delay
    - Removes word features from intelligibility transition
    - Focuses on SUT effects only
    - Smaller intelligibility transition region
- For aggregated words
  - t_0 average time to reach 50% intelligibility point
  - λ describes average intelligibility transition time
    - No word effects included
    - SUT dealing with different stimuli

Uncertainty

- Let $C = \ln\left(\frac{1-\alpha}{\alpha}\right)$
- Uncertainty for single, uncorrected access delay estimate

$$\mathrm{Var}(\tau(\alpha)) = C^2\mathrm{Var}(\lambda) + \mathrm{Var}(t_0) + 2C \cdot \mathrm{Cov}(\lambda, t_0)$$

- Propagation of uncertainty from linear combinations
  - Note that parameters for PTT and SUT curves are independent

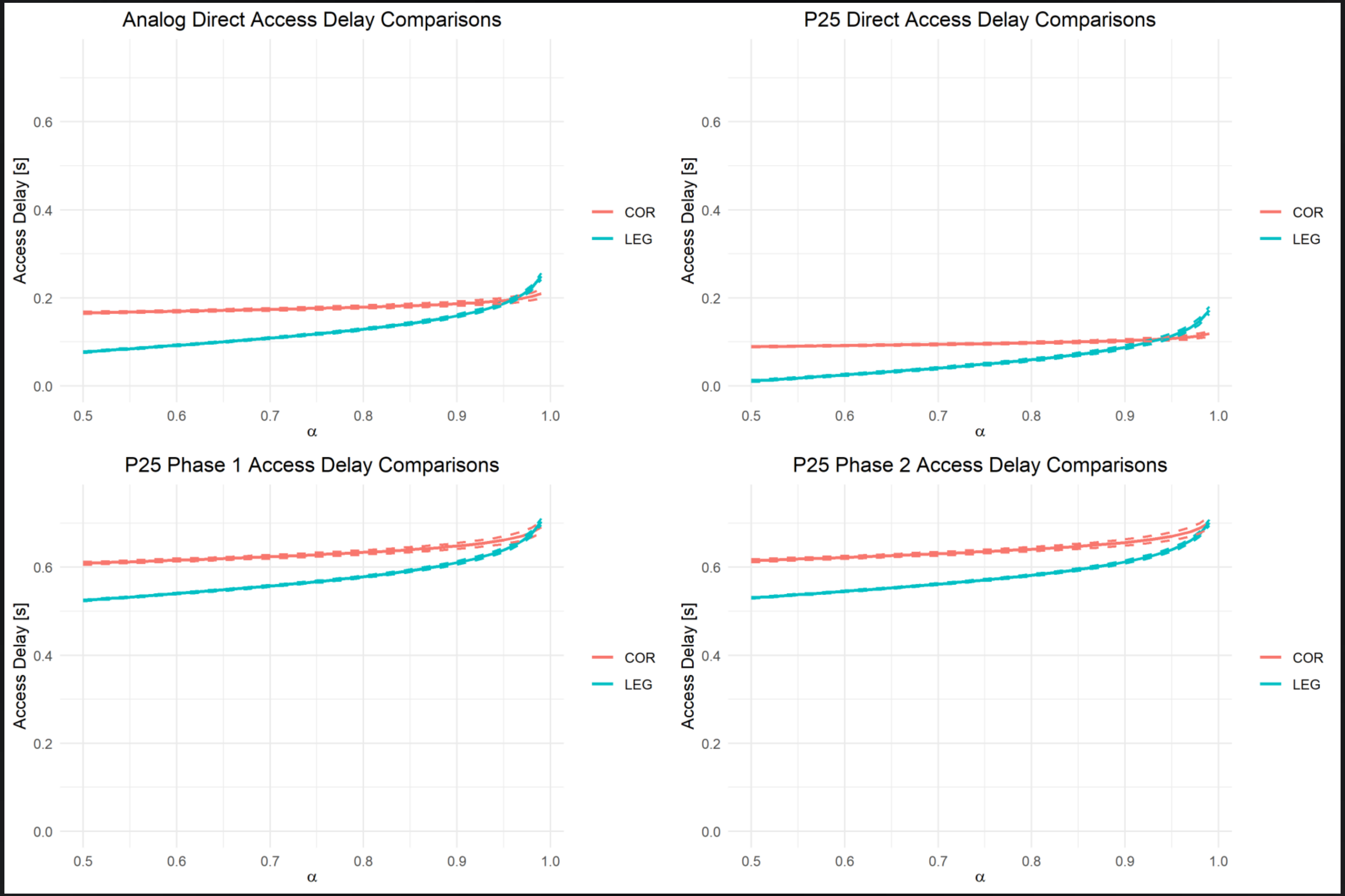
$$\mathrm{Var}(\tau_{\mathrm{sys}}(\alpha)) = \frac{1}{N^2}\left(C^2\sum_{j=1}^{N}\mathrm{Var}(\lambda_{j,\mathrm{C}}) + \sum_{j=1}^{N}\mathrm{Var}(t_{0,j,\mathrm{C}}) + 2C\sum_{j=1}^{N}\mathrm{Cov}(\lambda_{j,\mathrm{C}}, t_{0,j,\mathrm{C}})\right)$$

Impact on Measurement Results

- Access Delay increases across technologies
  - Difference between ~15ms to ~45ms
  - WUT are ‘hang’, ‘law’, ‘bed’, ‘not’
  - α fixed to 90%
  - 95% confidence interval included

Technology	New Access Delay $[ms]$	Old Access Delay $[ms]$	Difference $[ms]$
P25 Direct	103.0(99.6, 106.4)	87.9(83.4, 92.4)	15.14(9.49, 20.79)
P25 Phase 1	648.2(641.4, 655.1)	610.2(606.4, 614.0)	38.0(30.2, 45.8)
P25 Phase 2	656.4(649.4, 663.3)	612.2(608.7, 615.7)	44.1(36.3, 51.9)
Analog Direct	186.8(181.7, 191.9)	159.3(155.7, 162.8)	27.5(21.2, 33.7)

Impact on Measurement Results



Publications

- Mission Critical Voice Start of Word Correction for Access Delay Measurement System (2021)
  - Please see the link on the conference website
- Mission Critical Voice Quality of Experience Access Time Measurement Method Addendum (2020)
- Mission Critical Voice Quality of Experience Access Time Measurement Methods (2019)

Related PSCR 2021 Sessions

- Mission Critical Voice Quality of Experience Measurement Methods Overview
- Measuring the Probability of Successful Delivery: a QoE Based Approach
- Designing Remote Listening Experiments for the Partially Muted Word Impairment
- Optimal Transmit Volume Conditions for MCV QoE Measurement Systems
- QoE Software and Hardware Packaging
- QUARC: Quality Under Adjustable Realistic Conditions for Communication Systems
- Lab from Home: Distributed QoE Testing for Mission Critical Voice

Thank You

- Questions? Come to our Q&A!

THANK YOU

#PSCR2021 • PSCR.GOV