

FCC Telecommunications Relay Services Project

IP CTS Device Recurring Testing:

Prospective IP CTS Provider (MachineGenius) Assessment

April 29, 2020



Summary

- The National Test Lab (NTL) assessed Olelo, a software product developed by MachineGenius for its proposed Internet Protocol Captioned Telephone Service (IP CTS), which will use an automated speech recognition (ASR) platform without communications assistants (CAs). Specifically, the NTL assessed Olelo's Word Error Rate (WER) and caption delay compared to current IP CTS providers and ASR engines.
- Previously, the NTL tested the telephone captioning performance of IP CTS providers and general-purpose ASR engines using the same four call scenarios as in the Olelo test. Those test results indicated that:
 - The aggregated ASR engine results were faster (lower caption delay) than the aggregated IP CTS provider results.
 - The aggregated ASR engine WER was comparable to the aggregated IP CTS provider WER.
- The NTL compared Olelo's WER and caption delay test results to those of current IP CTS providers and ASR engines for the same call scenarios.
 - Olelo had lower (better) WER than the aggregated IP CTS provider and aggregated ASR engine results in all scenarios.
 - Olelo had lower (better) caption delay than the aggregated IP CTS provider and aggregated ASR engine results in all scenarios

Cross-Scenario Comparison of WER

- The table below shows aggregate WER summarized for existing IP CTS providers (dark blue), NTL-tested ASR engines (gray), and a prospective caption provider (MachineGenius) which uses an ASR engine (light blue)
 - For IP CTS providers, the data shown in the column with the dark blue header is aggregated (average of the median) WER for test calls to each Provider (five Providers)
 - For ASR engines, the data shown in the column with the gray header is the aggregated (average of the median) WER for test calls to both ASR engines
 - For MachineGenius, the data shown in the column with the light blue header is the median WER for the Olelo application

Word Error Rate (percent) by scenario and IP CTS provider / ASR engine – lower is better

| Scenario | Aggregated IP CTS Providers (revoiced / stenography) | Aggregated ASR | MachineGenius (ASR) |
|--------------------|--|----------------|---------------------|
| Long Call | 14.7 | 13.7 | 3.9 |
| Noise Cancelling | 18.1 | 47.3 | 8.2 |
| Professional Nurse | 8.9 | 8.1 | 2.7 |
| Lab Results | 19.5 | 22.3 | 5.5 |

Cross-Scenario Comparison of Caption Delay

- The table below shows aggregate caption delay summarized for existing IP CTS providers (dark blue), NTL-tested ASR engines (gray), and a prospective caption provider (Olelo) which uses an ASR engine (light blue)
 - For IP CTS providers, the data shown in the column with the dark blue header is the aggregated (average of the median) caption delay for test calls to each Provider (five Providers)
 - For ASR engines, the data shown in the column with the gray header is the aggregated (average of the median) caption delay for test calls to both ASR engines
 - For MachineGenius, the data shown in the column with the light blue header is the median caption delay for the Olelo application

Caption Delay (seconds) by scenario and IP CTS provider / ASR engine – lower is better

| Scenario | Aggregated IP CTS Providers (revoiced / stenography) | Aggregated ASR | MachineGenius (ASR) |
|--------------------|---|-------------------|---------------------|
| Long Call | 6.2 | 1.6 | 1.4 |
| Noise Cancelling | 4.8 | 1.4 | 1.1 |
| Professional Nurse | 5.3 | 1.8 | 1.3 |
| Lab Results | 6.8 | 1.6 | 1.1 |



Backup Slides – Scenario Details

Scenarios

- The NTL has been performing IP CTS device testing, and evolving IP CTS testing processes, since 2015. The testing represented in this deck incorporates lessons learned during previous testing and feedback from IP CTS providers, consumer groups, and the FCC Disability Advisory Committee. Based on this feedback these scenarios include:
 - Natural conversation including disfluencies and repeated words
 - Varying conversation lengths
 - Male/female voices
 - Varying background noise
 - Varying speech rates
 - Casual and formal conversations
- The NTL also performs other testing, including IP CTS usability testing. Usability test results are not included in this deck.

Long Call Scenario

**First used in IP CTS provider and ASR engine tests
conducted April 2018 – June 2018**

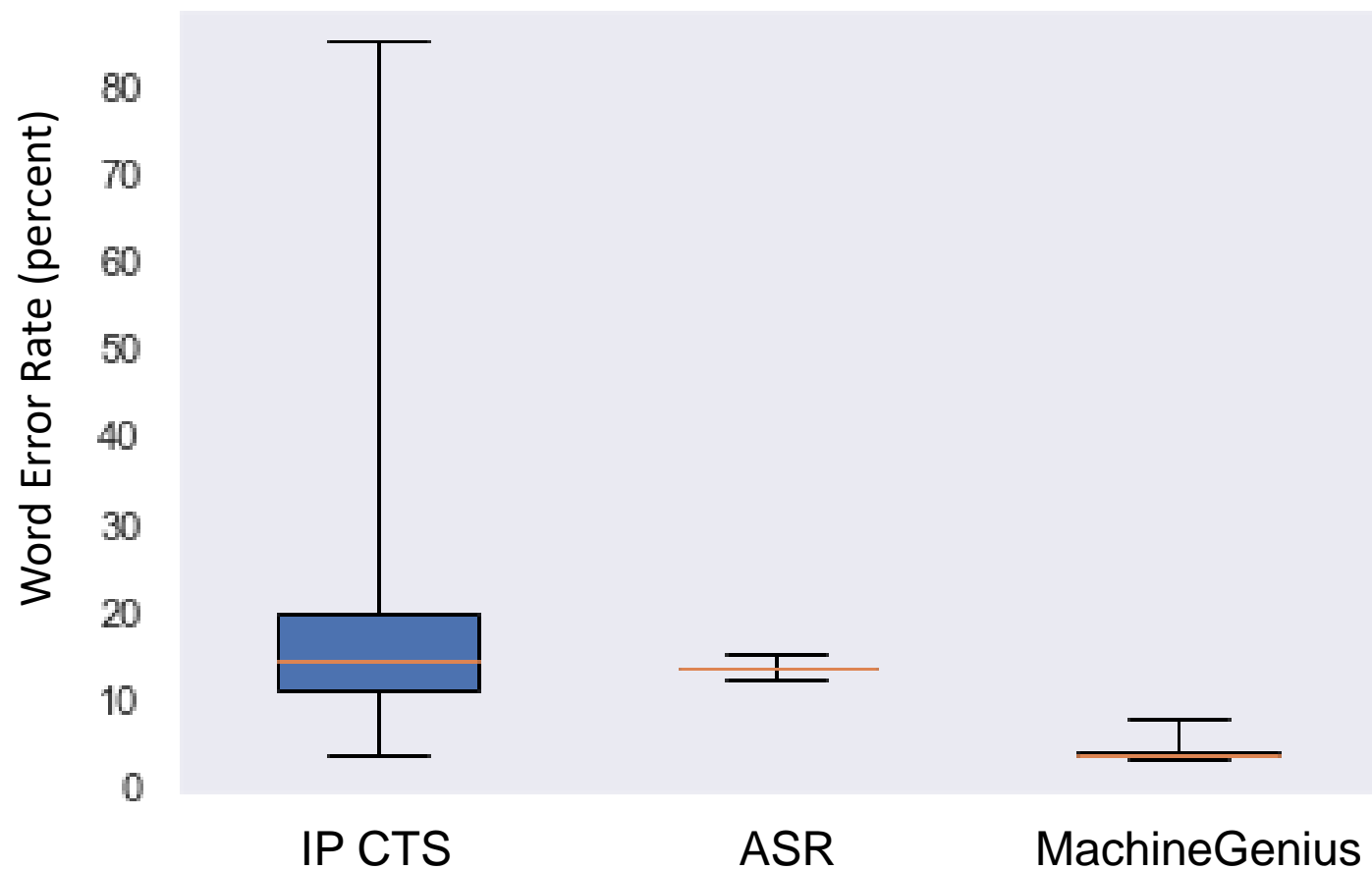
Long Call Scenario Summary

- The Long Call scenario is an 8:45 minute call that incorporates a female voice with rapid speech and varying volume levels.
- Timeframes - The NTL tested:
 - IP CTS providers and ASR engines between April 2018 – June 2018
 - ASR engines using NTL-developed tools were retested in November 2019. Latest results are shown.
 - MachineGenius (ASR) in March 2020

In this scenario,

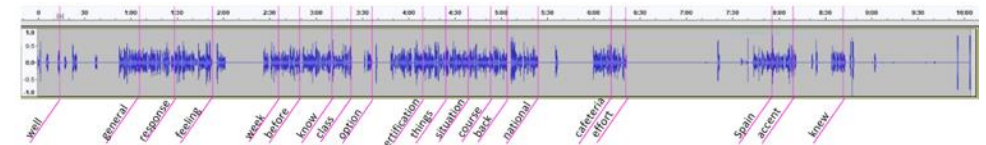
- *Olelo median WER (3.9 percent) was considerably lower (better) than aggregated WER for ASR engines and IP CTS providers*
- *Olelo median caption delay (1.4 seconds) was considerably lower (better) than aggregated caption delay for IP CTS providers*

Long Call Scenario – WER Comparison



| Service | Median WER |
|-------------------|------------|
| Aggregated IP CTS | 14.7% |
| Aggregated ASR | 13.7% |
| MachineGenius | 3.9% |

Long Call Scenario - Transcript



Words in **bold** are used for delay calculations

Hello? Hey Jim, I'm **well**. How are you? He did exceptionally well. Very, very intriguing. Very intriguing. Making some great strides. Understandably so. So, this week we're actually going to be taking what we were supposed to be discussing last week for the expressive portion. So, we're going to be learning greetings like, "Oh hi. How are you?" And then your **general** responses like we do in spoken English. So, I'm well, or I'm good, or I'm fine, or I'm great, or fantastic. Whatever, you know, the mood is that you're feeling. Uh, and then also the important part of that is attaching the facial expressions that come along with those types of **responses**. So, for example for fine, you can sign fine in two different ways. You can sign it in a way that is super sarcastic, "Or, you know, I'm fine. Or you can sign it in a way that says, "Oh, you know, I'm fine." So how you sign it along with what is actually signed is a - a huge part of somebody being able to let you know how they're **feeling**. Yes, exactly. Facial expressions can be equivalent to voice inflection.

Thank you. I really appreciate that. It's been, it's been quite a journey already and we're only, well, this will be the third **week**, so about three weeks in and people have come out of the woodwork left and right sharing their stories with me. Sharing personal stories with me nonetheless about why they're learning sign language. How they learned sign language **before**. And it's - it's been a humbling experience because not only are people sharing this personal thing with me but it's also just continuing to give me that push of passion that I have and continued immersion to just, you know, educate. Educate everybody on what I **know**. Spread, spread the wealth if you will. And uh, it's a really, really, incredible experience for my students to have such great feedback and really take, take well to the **class**. That's the best part about having a Skype **option**. Oh, I know.

It is going well. It is quite the process. Um, not sure if we've had the conversation before but there's a two-parter for interpreter certification. So right now, I have state screening, which is basically the same thing that I needed to do for my state **certification**. So, it's a two-parter. Most interpreting-type assessments are two-parters.

We have a written exam which is essentially testing our knowledge about interpreting, about our Code of Ethics, and all sorts of **things**. And then you have to pass that to go on to the next stage, or to the next test. And then the next test is the assessment of my ability to interpret, and if I interpret well, and what I would do in an ethical **situation**, and how I would solve that, you know, in an ethical manner, and certain things like that. And, uh, since I've already done the process before for the state level, I think it should go well for the national certification. But of **course**, with any kind of test-taking, it's - there's some anxiety that goes behind that. But I take that in a couple of weeks and then I have to wait a whopping 90 days to get those results **back**. Yeah, it was the same. Yeah. That was the best part about the written exam was that the results were instantaneous. But I had to wait about three, three and a half months for my state screening, so I wasn't surprised to see that I'd have to wait around the same time for my **Nationals**. That is, that is so true.

Well, I must commend you for what you have done, Jim. Because I see the interactions that you have with Evan and I see that you are genuinely trying to learn. And even when we have conversations in the **cafeteria**, there are things that you are picking up on that others are not. And you're also asking, "Oh, what's the sign for this?" So, you're putting forth a lot more **effort** than you may even realize. Hey, better than zero.

It's funny that you say that though. Because I was out with some friends this past Friday or Saturday. I can't remember. And I was chatting with a woman from **Spain**. And bear in mind my foreign language in high school and college was ASL, so I - I know very, very little Spanish. And I asked her, I said, "Oh my goodness. Where are you from? Because, you know, I can hear it in your **accent**." And she said, "Oh, I'm from Spain." That's five more words than I did. And I didn't even know that made, that meant "Thank you very much." I only **know** that saying from the Styx song. Good to know. You still there, Jim?

“Long Call” Scenario Audio Characteristics

- The “Long Call” scenario incorporates a female voice with rapid speech and varying volume levels.
- The list below provides the calculated source audio characteristics of the “Long Call” scenario and provides the basis for assessing the “difficulty” of a particular scenario.
 - Speaker gender: Female
 - Word Count: 813 words
 - Speech Rate: 92.97 words per minute
 - Speaking Ratio: 47 Percent
 - Reading Grade Level: 5.0
 - Call Duration: 8:45 minutes
 - Background noise: low
 - Speaker volume level: variable (medium to high)
 - Speaker clarity: variable (medium to high)

Noise Cancelling Scenario

First used in IP CTS provider and ASR engine tests
conducted June 2018 – August 2018

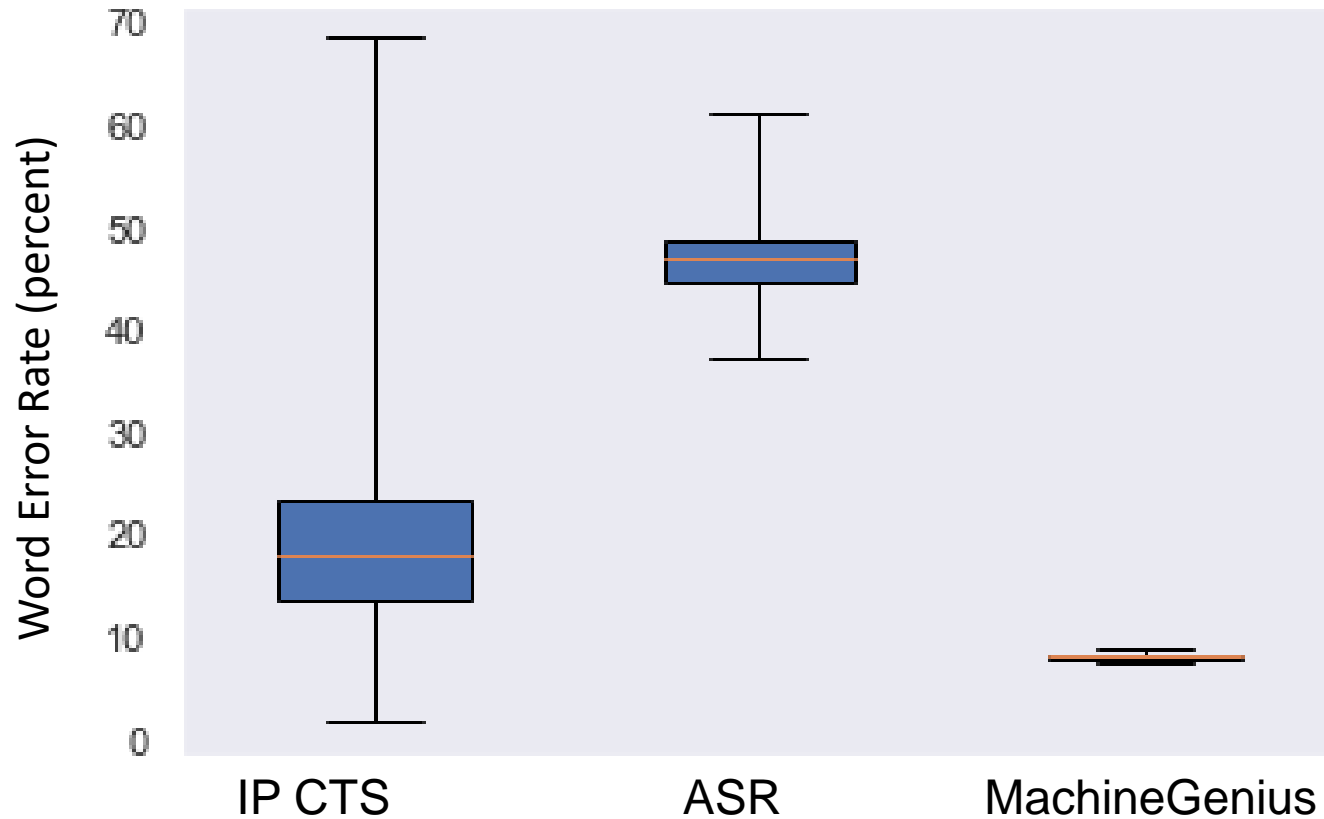
Noise Cancelling Scenario Summary

- The “Noise Cancelling” scenario is a 4:10 call that incorporates a high-pitched female voice with rapid speech, varying volume levels, and significant background noise.
- Timeframes - The NTL tested:
 - IP CTS providers and ASR engines between June 2018 – August 2018
 - ASR engines using NTL-developed tools were retested in December 2019. Latest results are shown.
 - MachineGenius (ASR) in March 2020

In this scenario,

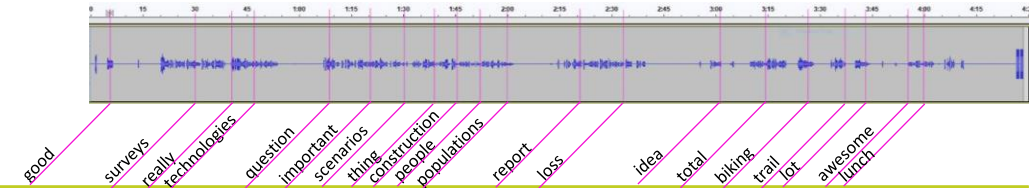
- *Olelo median WER (8.2 percent) was considerably lower (better) than WER for aggregated ASR engines and IP CTS providers*
- *Olelo median caption delay (1.1 seconds) was considerably lower (better) than aggregated caption for IP CTS providers*

Noise Cancelling Scenario – WER Comparison



| Service | Median WER |
|-------------------|------------|
| Aggregated IP CTS | 18.1% |
| Aggregated ASR | 47.3% |
| MachineGenius | 8.2% |

Noise Cancelling Scenario Transcript



Words in **bold** are used for delay calculations

Hello? Hi, Jim. I'm **good**. How are you? Sure. So, we're going over to the HLA convention next (month) or in a couple of weeks in Minneapolis, and we're administering a couple of **surveys**, and we're aiming at least 30 participants for the surveys, but ideally, we would like as many as possible. We **really** want to get -- yeah. We really want to get a representative sample of who might be using these **technologies** and what technologies they're using so we can get a better understanding of what the general population's using. That's - That's a good **question**. It's not always the same, um, but for our purposes, um, we ask a lot of demographic information. That's really **important**. We want to make sure that we're getting people who have various levels of hearing loss and people who utilize different technologies for different **scenarios**. Um, and different ages as much as possible. Um, but we know that it's not always the easiest **thing**, um, to do when you're at a certain conference -- sorry. I was going by some **construction**. It's not always the easiest thing when we go to a conference that really tailors itself to one group of **people**, but we'll do the best we can. And, um, from there, we'll be able to figure out what other groups of, um, people from what other **populations** we need to talk with, so...

Sure. So, our surveys this time around are 100% self-**report**. So, we're asking the participants if they know like what decibel, um, of hearing **loss** they have. That's one of the questions that they can answer, the degree to which they have their hearing loss. We also ask them how they, um, how they identify with their hearing loss. Yes. Yeah. I think that sounds like a really cool **idea**. (Garbled) In the future. Um, About 5 minutes away. It's about 30 minutes **total**. Even though I live about 3 miles away, it takes me 30 minutes door to door sometimes. Oh, my God. I know. Sometimes I think about **biking** in, but I never do it. Yeah. It's tough. Out in Reston, there's a, um, bike **trail**. Yeah. I live a mile off of it, so I hop on that a **lot** and ride out to Reston. Yeah. Alright. **Awesome**. Um. I'll be in -- I'll be in and out for the rest of the morning through **lunch**, but, uh, is it okay if I pop in this afternoon? Oh, awesome. Okay, great. Thanks, Jim. Alright. Bye.

Noise Cancelling Scenario Audio Characteristics

- The “Noise Cancelling” scenario incorporates a high-pitched female voice with rapid speech, varying volume levels, and significant background noise.
- The list below provides the calculated source audio characteristics of the “Noise Cancelling” scenario and provides the basis for assessing the “difficulty” of a particular scenario.
 - Speaker gender: Female
 - Word Count: 420 words
 - Speech Rate: 100.71 words per minute
 - Speaking Ratio: 58 Percent
 - Reading Grade Level: 4.8
 - Call Duration: 4:10 minutes
 - Background noise: high
 - Speaker volume level: variable (medium to high)
 - Speaker clarity: variable (medium to high)

Professional Nurse Scenario

First used in IP CTS provider and ASR engine tests
conducted 12/2018 – 04/2019

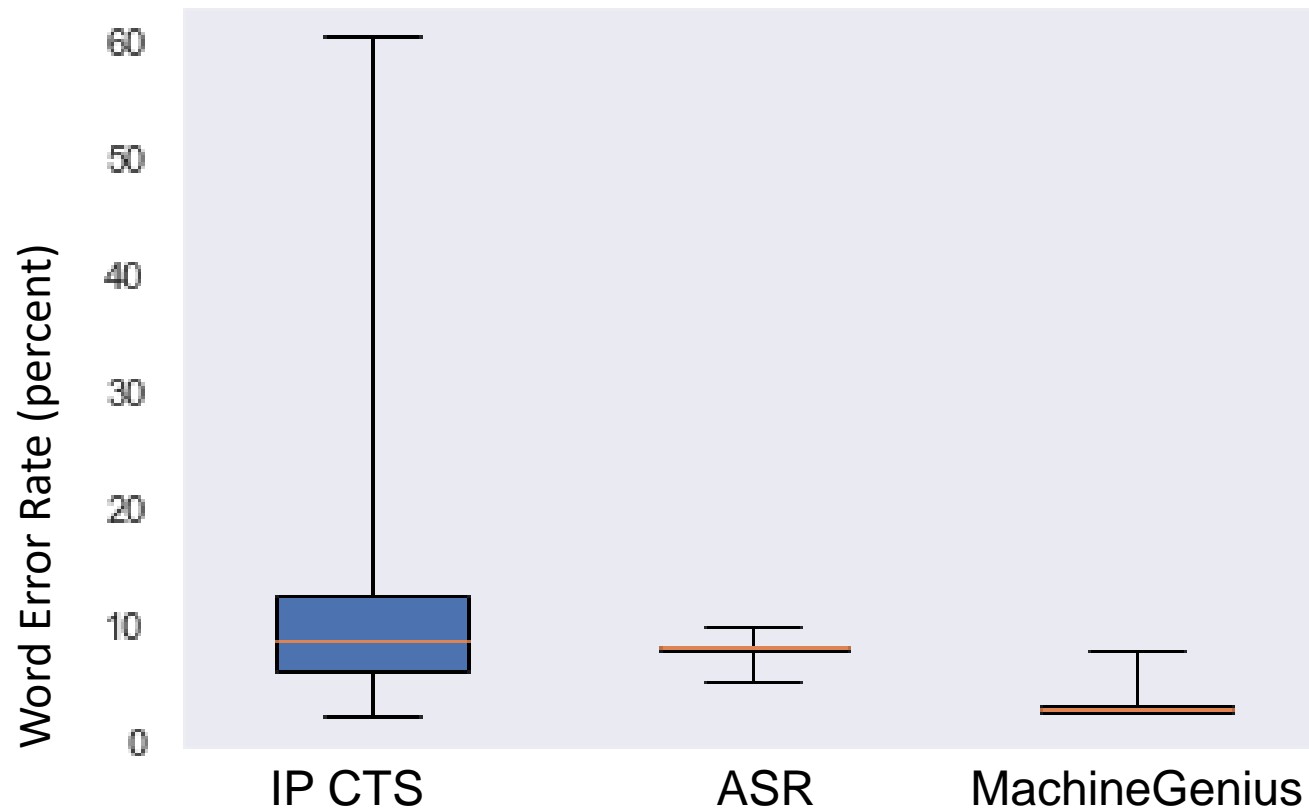
Professional Nurse Scenario Summary

- The “Professional Nurse” scenario is a 5:45 minute call that depicts a female nurse with minimal accent describing the Shingles vaccine. The conversation includes medical terminology as well as medicine names. This scenario was recorded in a “clean” office environment (i.e., no background noise, fixed volume).
- Timeframes - The NTL tested:
 - IP CTS providers and ASR engines between December 2018 – April 2019
 - MachineGenius (ASR) in March 2020

In this scenario,

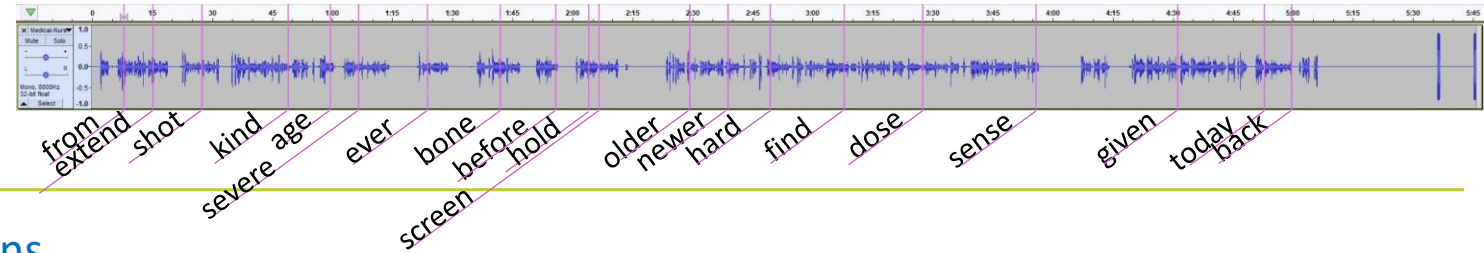
- *Olelo median WER (2.7 percent) was considerably lower (better) than aggregated WER for ASR engines and IP CTS providers*
- *Olelo median caption delay (1.3 seconds) was considerably lower (better) than aggregated caption delay for IP CTS providers*

Professional Nurse Scenario – WER Comparison



| Service | Median WER |
|-------------------|------------|
| Aggregated IP CTS | 8.9% |
| Aggregated ASR | 8.1% |
| MachineGenius | 2.7% |

“Professional Nurse” Transcript



Words in **bold** are used for delay calculations

Hi. Is this Mr. Taylor? Mr. Taylor, this is Ashley **from** Dr. Thomas' office. Um. You had given us a call and I wanted to get back with you. Before we **extend** this conversation, could you tell me your, uh, date of birth?

Um, Dr. Thomas has explained to me that he wanted you to get the **shot** for Shingles, also known as Herpes Zoster. Is that correct?

Good. Alright. So, I'm calling to go over the vaccines with you, available vaccines with you. The first thing I'd like to do is to ask you, um, a question from the CDC. They **kind** of go over um, whether or not you need this vaccine. Is that ok?

Alright, so are you 60 years of **age** or older? Have you ever had a life threatening or **severe** allergic reaction to gelatin, the antibiotic Neomycin, or any other vaccine component?

Um. Have you **ever** had treatment for cancer such as radiation or chemotherapy?

Ok. Um, and 2 more. Uh, History of cancer affecting the **bone** marrow or lymphatic system such as leukemia or lymphoma?

Ok, and the last one is you've have you ever had any vaccine for shingles **before**?

Ok. Alright so if you could just **hold** on for one second, I want to get to the **screen** that will help me explain the 2 vaccines that are available to you. Is that ok? Thank you.

Ok. I'm back. Um. There are 2 vaccines currently on the market. The **older** vaccine called Zostavax is for those, um, adults over 50 years of age. It's been out for quite a while. There's a **newer** vaccine on the market called Shingrix. Um. It, uh, it is **hard** to get. Uh. Very hard to get because it has become very popular. It's the one that the CDC recommends. Um. What I would like to do is to go ahead and send you a prescription for Shingrix and recommend that you contact your pharmacy before you go there to **find** out if they have it. If they do have it, then ask them to put that 2nd **dose** aside. It's administered in 2 doses. They have to be 6 months apart. If you have a problem getting that vaccine, and the pharmacy says they're not going to have it or they can't put that 2nd dose aside then give me a call and we can go back and send you a prescription for the Zostavax. That way at least you will be protected. Um, and then we could always go ahead once, uh, GSK or Glaxo-Smith-Kline puts more Shingrix available on the market, you can go ahead and get that later on. Uh, because Dr. Thomas felt it was important that you have that. Um. Does that seem to make **sense** to you?

Are you ok with me sending a prescription for the, um, vaccine? Ok. Do you have any questions?

It's liable to make you feel a little tired. Of course, like with any vaccine you may get some redness, some pain, a little swelling at the site where the shot was **given** but both of them are relatively safe. Um, those symptoms can be taken care of with [Acetaminophen](#) or Tylenol or um, you know, over the counter Motrin, Ibuprofen, something like that. Um, side effects are very short-lived.

Alright, so I will get that out to you **today**. And if you have any more problems or questions once you contact the pharmacy, don't hesitate to give me a call **back**.

Alright. Thank you, Mr. Taylor. You have a good day. Alrighty. Bye-bye.

“Professional Nurse” Audio Characteristics

- The “Professional Nurse” scenario depicts a female nurse with minimal accent describing the Shingles vaccine. The conversation includes medical terminology as well as medicine names. This scenario was recorded in a “clean” office environment (i.e., no background noise, fixed volume).
- The list below provides the calculated source audio characteristics of the “Professional Nurse” scenario and provides the basis for assessing the “difficulty” of a scenario.
 - Speaker gender: Female
 - Word Count: 600 words
 - Speech Rate: 118.38 words per minute
 - Speaking Ratio: 70 Percent
 - Reading Grade Level: 4.1
 - Call Duration: 5:45 minutes
 - Background noise: low
 - Speaker volume level: medium
 - Speaker clarity: medium to high

Lab Results Scenario

First used in IP CTS provider and ASR engine tests
conducted 05/2019 – 06/2019

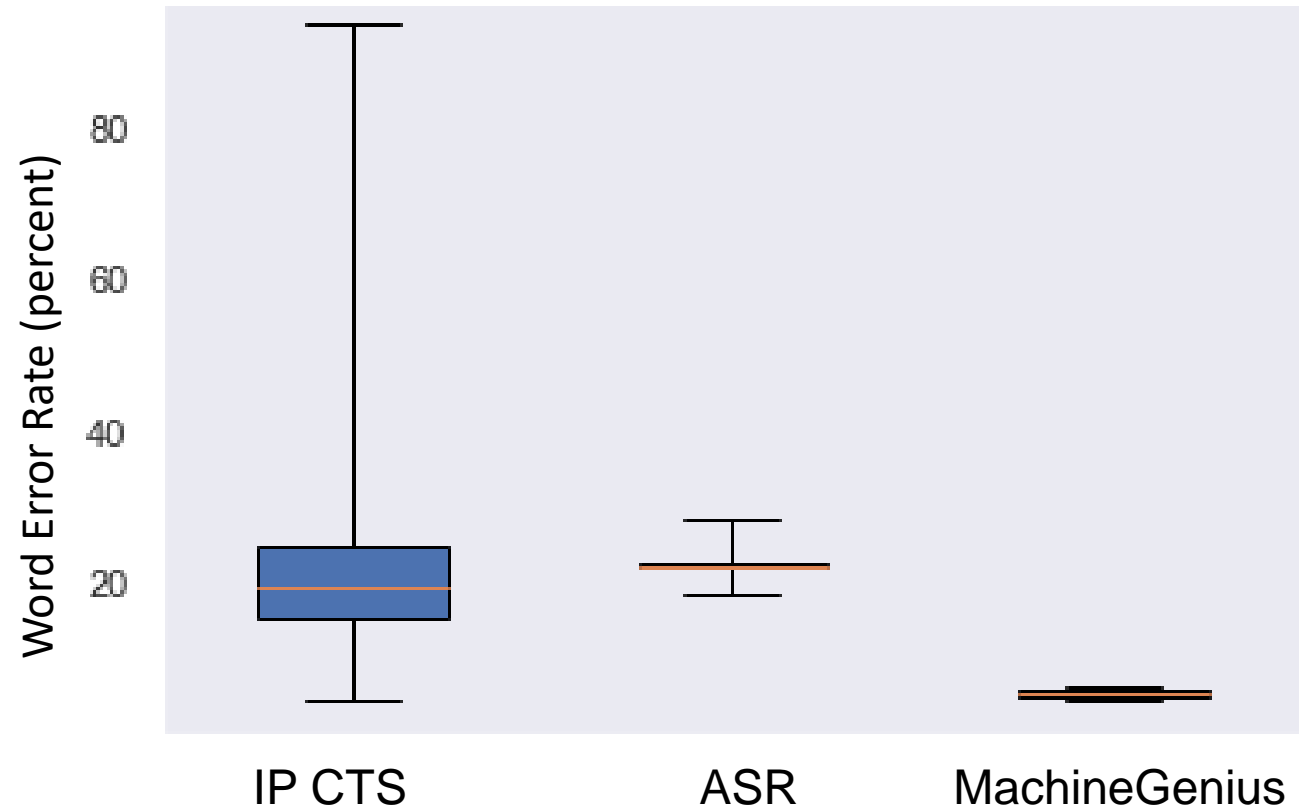
Lab Results Scenario Summary

- The “Lab Results” scenario is a 3:46 call that depicts a male voice with minimal accent and rapid speech describing medical lab results. The conversation includes medical terminology as well as medicine names. This scenario was recorded in a “clean” office environment (i.e., no background noise, variable volume).
- Timeframes - The NTL tested:
 - IP CTS providers and ASR engines between May 2019 – June 2019
 - MachineGenius (ASR) in March 2020

In this scenario,

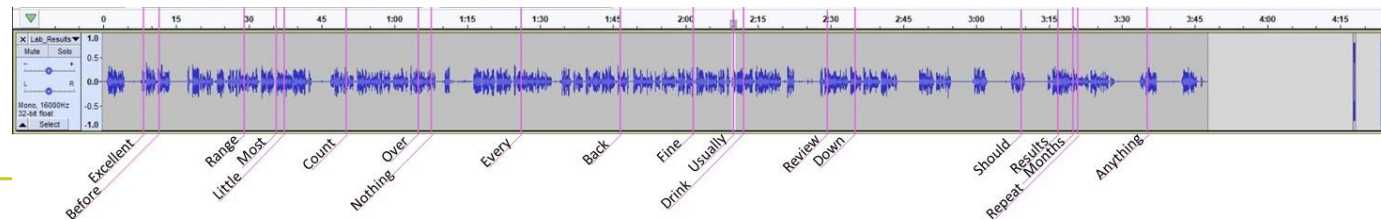
- *Olelo median WER (5.5 percent) was considerably lower (better) than aggregated WER for ASR engines and IP CTS providers*
- *Olelo median caption delay (1.1 seconds) was considerably lower (better) than aggregated caption delay for IP CTS providers*

Lab Results Call Scenario – WER Comparison



| Service | Median WER |
|-------------------|------------|
| Aggregated IP CTS | 19.5% |
| Aggregated ASR | 22.3% |
| MachineGenius | 5.5% |

“Lab Results” Transcript



Words in **bold** are used for delay calculations

Hi. This is Matthew from Springfield Medical Group. I'm calling for Sharon Miller. **Excellent**. I have some lab results I'd like to go over with you. Uh, **before** we start, can you confirm your date of birth for me? Thank You. Dr. Perez had us run 5 panels for you and I just want to walk through the results of those. The first is a metabolic panel. The results, uh, came back, uh, everything was in **range** except for the creatinine levels. Uh, that's a - a measure of kidney function. **Most** likely that simply means you were a **little** dehydrated on the day of the test. We'll want to follow up on that. I'll - I'll talk about follow-ups at the end of the call though.

The 2nd one was a, uh, complete blood **count** - a CBC. Uh, everything in there was well within ranges except for the lymphocytes and monocytes uh, which were a little high. That's generally an indication of an infection, um, uh, just barely **over** the lines though. Probably something you're just - just getting over. **Nothing** to worry about.

Any questions so far?

The 3rd panel was uh, TSH the Thyroid Stimulating Hormone. Uh, given your family history of Hashimoto's disease, um, we check that one **every** time and yours is at 2.5 which is just in the middle of the range, so your thyroid is doing everything it's supposed to.

Fourth panel was cardiac markers. They look for cardiac troponin um as a - an indicator of uh, stroke or - or uh, myocardial infarction. Uh, you came **back** completely negative which is good.

The last was the uh, peripheral capillary oxygen saturation – SPO2. That was at 97 percent which is again very good. So, all of your - your uh, blood tests came back just **fine**. Um, do watch your hydration. The – the uh, that one at the beginning there – the creatinine levels um **usually** that just means you haven't had enough to **drink** that day which - which is not uncommon. If it's continually out of range uh, then we'll have to do some more testing to try to identify why.

Any questions about the test results?

OK. Uh, let me **review** medications make sure that uh, what you're taking and what we think you're taking match. I have you **down** as taking Azelastine, 137 micrograms. That bottle may say Astepro. That's the uh, brand name. And Diltiazem, 120 milligrams. Uh, Again, that - the uh, - it might say Cardizem.

And uh, Aspirin, 81 milligrams.

Fantastic. Are you taking any other meds we **should** be aware of?

OK. Uh, because of the **results** we talked about, uh, we do want to bring you back in in six **months** rather than a year to **repeat** some of the bloodwork and just do a general checkup. Uh, if you can call when it's convenient and schedule an appointment with Dr. Perez, we'd appreciate it.

Fantastic. Is there **anything** else I can help you with today?

Alright, well thank you for your time and uh, we'll talk to you soon. Bye.

“Lab Results” Audio Characteristics

- The “Lab Results” scenario is a 3:46 call that depicts a male voice with minimal accent and rapid speech describing medical lab results. The conversation includes medical terminology as well as medicine names. This scenario was recorded in a “clean” office environment (i.e., no background noise, variable volume).
- The list below provides the calculated source audio characteristics of the “Lab Results” scenario and provides the basis for assessing the “difficulty” of a scenario.
 - Speaker gender: Male
 - Word Count: 509 words
 - Speech Rate: 135.28 words per minute
 - Speaking Ratio: 67 Percent
 - Reading Grade Level: 5.3
 - Call Duration: 3:46 minutes
 - Background noise: low
 - Speaker volume level: Variable - low to medium
 - Speaker clarity: Medium

Data Rights Notice

NOTICE

This presentation was produced for the U. S. Government under Contract Number 75FCMC18D0047, and is subject to Federal Acquisition Regulation Clause 52.227-14, Rights in Data-General.

No other use other than that granted to the U. S. Government, or to those acting on behalf of the U. S. Government under that Clause is authorized without the express written permission of The MITRE Corporation.

For further information, please contact The MITRE Corporation, Contracts Management Office, 7515 Colshire Drive, McLean, VA 22102-7539, (703) 983-6000.