WILKINSON) BARKER) KNAUER) LLP

1800 M STREET, NW
SUITE 800N
WASHINGTON, DC 20036
TEL 202.783.4141
FAX 202.783.5851
WWW.WBKLAW.COM
DAVID A. O'CONNOR
202.383.3429
DOCONNOR@WBKLAW.COM

December 19, 2018

VIA ELECTRONIC FILING (ECFS)

Marlene H. Dortch, Esq.
Secretary
Federal Communications Commission
445 Twelfth Street, SW
Washington, DC 20554

> RE: **EX PARTE PRESENTATION**
> *Misuse of Internet Protocol (IP) Captioned Telephone Service;*
> *Telecommunications Relay Services and Speech-to-Speech Services for*
> *Individuals with Hearing and Speech Disabilities*
> CG Docket Nos. 13-24, 03-123

Dear Ms. Dortch:

On December 17, 2018, representatives of Hamilton Relay, Inc. ("Hamilton") and CapTel Inc./Ultratec Inc. ("Ultratec") met with Commission staff from the Consumer & Governmental Affairs Bureau ("CGB") and the Office of the Managing Director ("OMD"). Participating on behalf of Hamilton were Dixie Ziegler, John Nelson (by telephone), Jeff Knighton, Rachel Wolkowitz (outside counsel) and the undersigned counsel (by telephone). Participating on behalf of Ultratec were Robert Engelke, Christopher Engelke, Kevin Colwell, and Jayne Turner. Commission staff present at the meeting were Karen Peltz Strauss, Eliot Greenwald, Michael Scott, and Robert Aldrich of CGB, and David Schmidt of OMD (by telephone).

During the meeting, the parties shared their initial findings from an extensive new study examining Automatic Speech Recognition ("ASR") and Communications Assistant ("CA") performance in the context of Internet Protocol Captioned Telephone Service ("IP CTS"). In particular, the study examines the relative importance of accuracy and delay (traditional proxies for "Functional Equivalence" in relay) on users' perceptions of captioning usefulness in real-world environments.[1] First, the research indicates that accuracy influences users' perceptions of

---

[1] The parties presented from the following slide deck, which summarizes Hamilton's and CapTel's methodology and findings.

delay and therefore poses an independent "first hurdle" before considering the dependent variable of delay. Secondly, the research shows that weighing accuracy over speed would better represent usefulness of captions and ensure greater functional equivalence for users. The double-blind study included over 6,500 calls, some of which were captioned by CAs and some of which were captioned by ASR.

Hamilton urges the Commission to consider the study's findings as it develops objective metrics for assessing IP CTS quality of service, as well as how and when to integrate ASR-only captions into the IP CTS program generally.[2] The parties welcome comment from the public on the study's initial findings, and believe that additional research is needed in this area.

This filing is made in accordance with Section 1.1206(b)(2)(iv) of the Commission's rules, 47 C.F.R. § 1.1206(b)(2)(iv). In the event that there are any questions concerning this matter, please contact the undersigned.

Respectfully submitted,

WILKINSON BARKER KNAUER, LLP

/s/ David A. O'Connor
Counsel for Hamilton Relay, Inc.

cc (via email):   Karen Peltz Strauss
                Eliot Greenwald
                Michael Scott
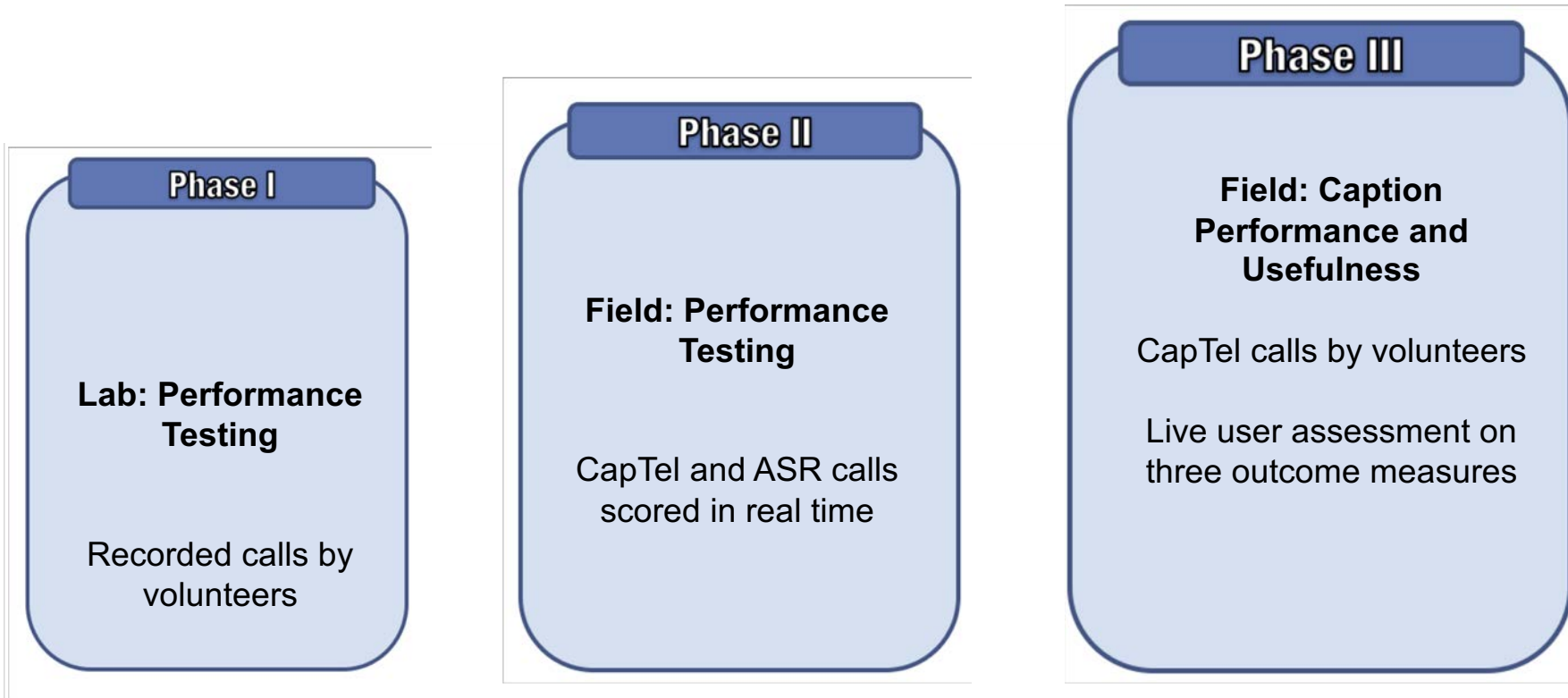   Robert Aldrich
                David Schmidt

Enclosure

---

[2] *See Misuse of Internet Protocol (IP) Captioned Telephone Service; Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, Report and Order, Declaratory Ruling, Further Notice of Proposed Rulemaking, and Notice of Inquiry, 33 FCC Rcd 5800 (2018), Petition for Reconsideration pending. Hamilton notes that both it and other parties have objected to the *Declaratory Ruling* on procedural and substantive grounds, and Hamilton believes that it is premature to authorize ASR-only IP CTS service before these objective metrics are implemented.

A Trial of Automated Speech Recognition for IP CTS Calls

Phase 3 Report

Three Phase Trial of ASR in IP CTS

Phase I

Lab: Performance Testing

Recorded calls by volunteers

Phase II

Field: Performance Testing

CapTel and ASR calls scored in real time

Phase III

Field: Caption Performance and Usefulness

CapTel calls by volunteers

Live user assessment on three outcome measures

Three Phase Trial of ASR in IP CTS

- Hamilton and CapTel conducted a three phase trial of ASR in the field of IP CTS.

- Testing was done in both laboratory and "real life" conditions.

- All tests were double-blind (neither users nor scoring agents were aware of which condition of which they were part: ASR or CA).

- All scoring was done with a proprietary, real-time scoring tool ensuring that reliable metrics could be collected and that **no call content was recorded**.

- No calls made from trial phones during the trial were submitted for reimbursement.

Functional Equivalence

Directly Observable Qualities

       Accuracy

       Delay

Not Directly Observable Qualities

       Experience

       Confidence

       Empowerment
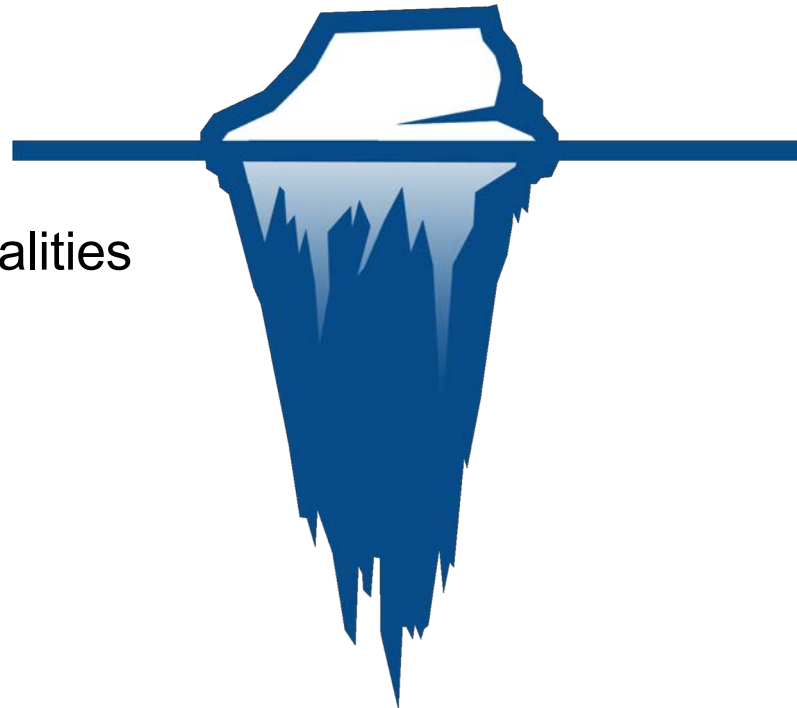
Functional Equivalence

- Accuracy and delay have long served as proxies for measuring "Functional Equivalence." However, no study to date has looked at the relative importance of accuracy and delay on users' perceptions of caption usefulness in real-world environments.

- Usefulness is a product of qualities that are both directly measurable (e.g. accuracy and delay) and not directly measurable (e.g. domains of user experience).

- Hamilton and CapTel conducted rigorous qualitative and quantitative testing to determine the relative importance of accuracy and speed on caption usefulness (functionality) in real-world environments.

Phase III: Performance and Usefulness

Phase III

Field: Caption Performance and Usefulness

CapTel calls by volunteers

Live user assessment on three outcome measures

2-week duration
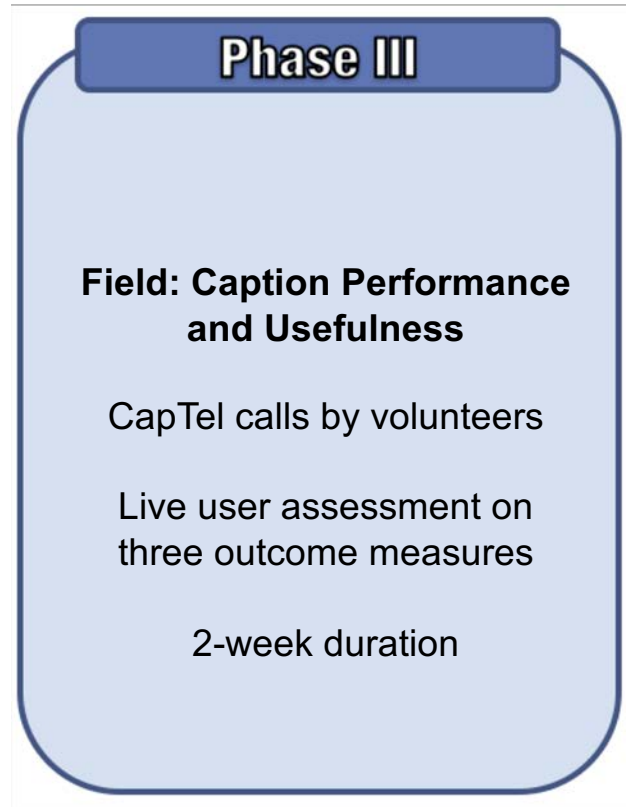
- 117 volunteer participants, ~6,500 calls

- Volunteers were all registered IP CTS users and used the same (familiar) hardware and interface at home/office for two weeks

- Volunteers were surveyed after each call on performance and usefulness of captions

- Volunteers were paid per questions answered (up to $100/wk)

- No calls throughout all phases of the trial were submitted for reimbursement

- 911 calls were directed to standard service

- Accuracy and Delay measured in real time (no call content recorded)

- **Double-Blind Design**

See what they say

CapTel

HAMILTON
captel

Three Phase Trial of ASR in IP CTS

- 117 volunteers made approximately 6,500 calls through the test services. Captions were generated by either CA or ASR. After every call, volunteers were asked to estimate the accuracy and delay of the call, and were surveyed across 3 areas of functionality and usefulness. Accuracy was measured in real time by a CapTel quality monitor (no call content was recorded). Neither the volunteer nor the quality monitor knew which system had produced the captions. Volunteers were paid up to $100/wk based on how many questions they answered.

- Volunteers were all CapTel 2400iBT users prior to enrolling in the testing and were given 2400iBT phones to use for the duration of the tests. Every volunteer used the phone in his/her home and/or office for a period of 2 weeks.*

- Call minutes were not submitted to the fund administrator for reimbursement.

*A small number of volunteers exited the study early

Participant Demographics

Gender

57 %
43 %

State

Education

High School Graduate

Post Secondary Diploma

Participant Demographics

Severe 53%

Moderate 41%

Mild 6%
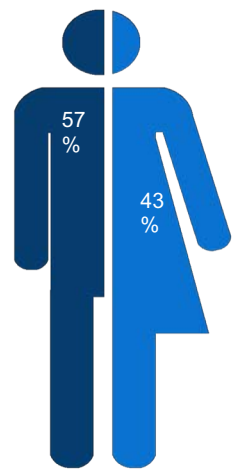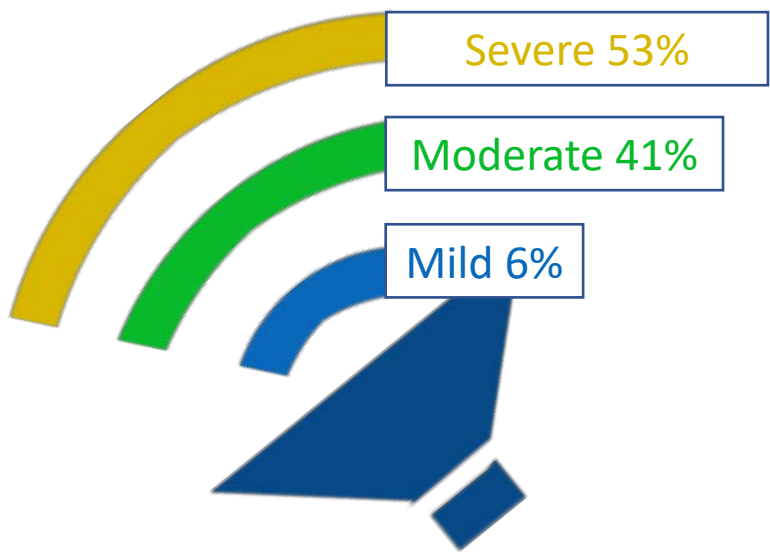
Self-Reported Level of Hearing Loss

Hearing Aid or Cochlear Implant Use

97% Yes

3% No
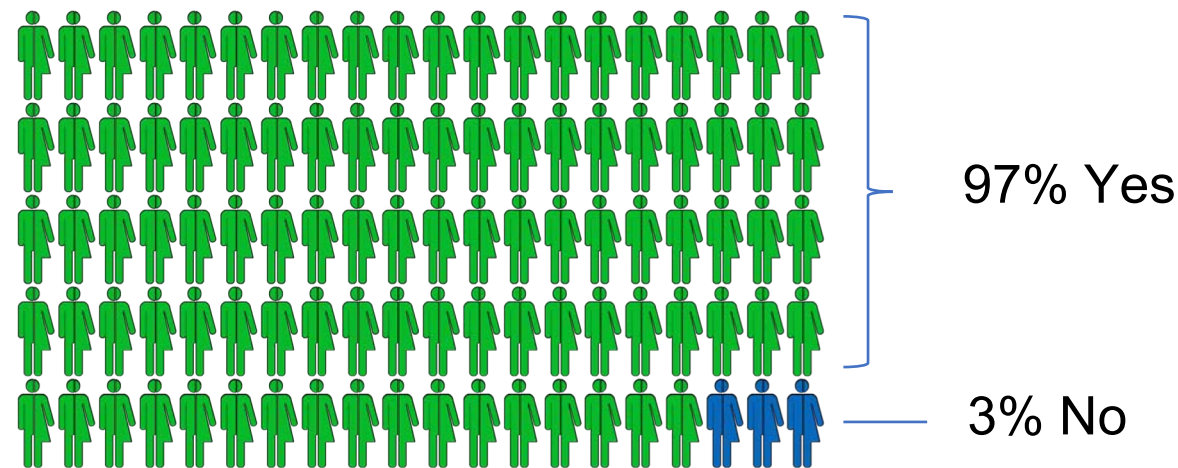
Volunteers Reflect CapTel User Demographics

- 117 volunteer participants from 31 States.

- 97% of volunteer participants reported using at least one hearing aid or cochlear implant.

- 94% of volunteer participants reported moderate to severe hearing loss.
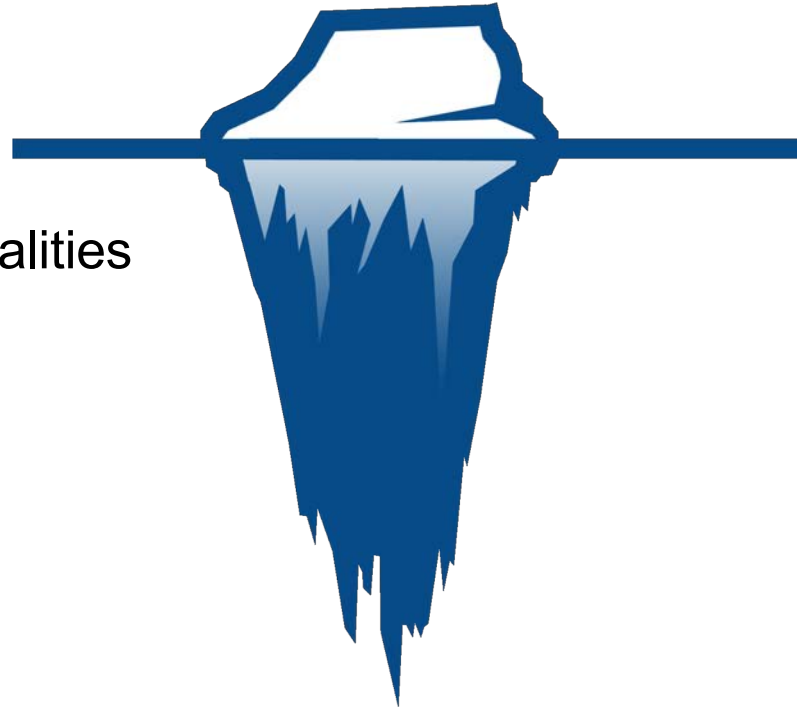
Functional Equivalence

Directly Observable Qualities
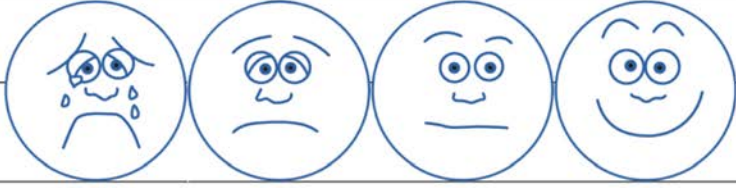
    Accuracy

    Delay

Not Directly Observable Qualities

    Experience

    Confidence

    Empowerment

Not Directly Observable Qualities – Usefulness

Latent Construct	Indicator Item	😢	😟	😐	😊
Experience with the call	1. Overall, how was the call flow?	Terrible	Not Good	Good	Great
	2. Overall, how satisfied were you with the call?	Not at all	Not very	Somewhat	Very satisfied
Confidence in Captions	3. How dependable are these captions?	Not at all	Not very	Somewhat	Extremely
	4. How useful are these captions?	Not at all	Not very	Somewhat	Extremely
	5. How reliable do you find the captions?	Not at all	Not very	Somewhat	Extremely
User Empowerment	6. Would you feel confident using this service for business calls?	No way	Not really	Probably	Absolutely
	7. How safe would you feel using this service in an emergency?	No way	Not really	Probably	Absolutely
	8. Does the technology allow you to feel connected to the other person?	No way	Not really	Kind of	Absolutely

CapTel

See what they say®

HAMILTON® captel

Examples of User Screen
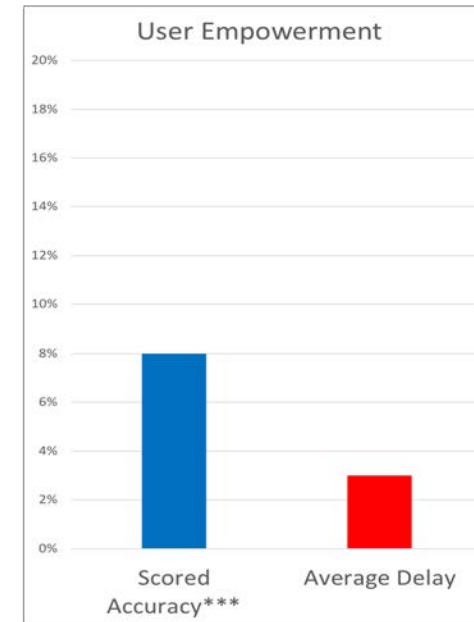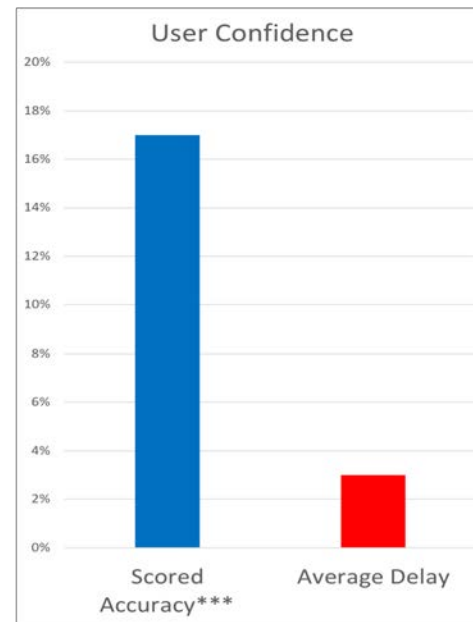
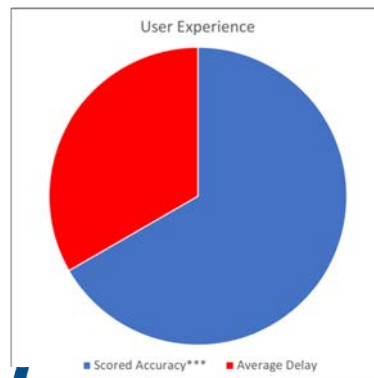# Not Directly Observable Qualities – Usefulness
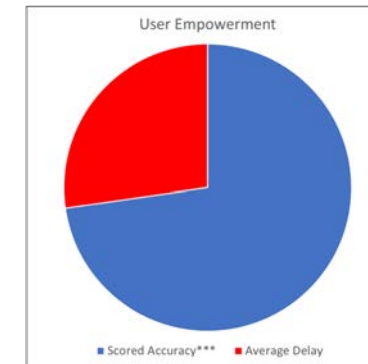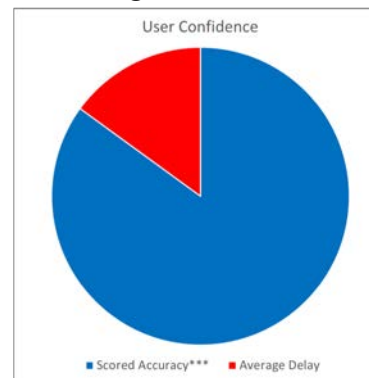
- Volunteers were given a survey after every call.
  - Estimate Accuracy
  - Estimate Delay
  - 4 "usefulness" Questions
    - Questions and question order randomized to avoid "order bias"
- Calls were evaluated in real-time (no content recorded).
  - Accuracy was assessed by a trained quality monitor
  - Average delay was computed based on final form of word
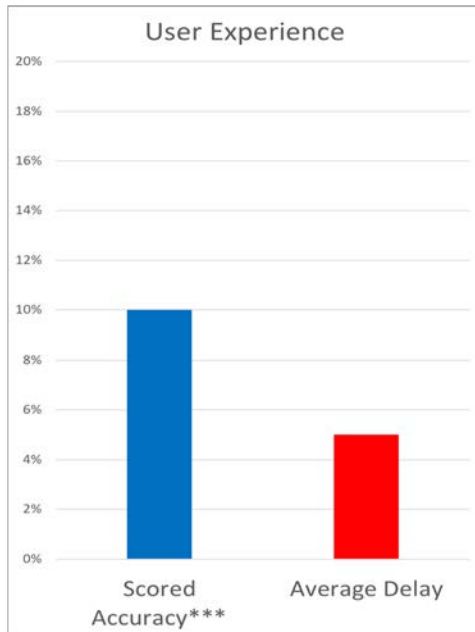
Relative Impact of Scored Accuracy and Delay*

Relative Percentage of Variance Explained

*Comparison of Standardized Regression Coefficients (Predictive Power)

Relative Impact on User Experience*

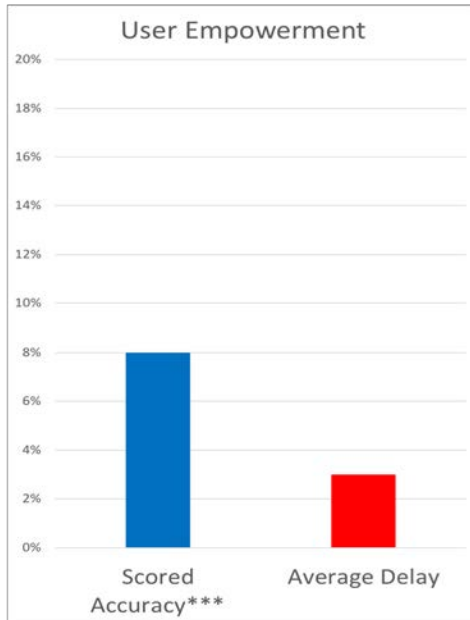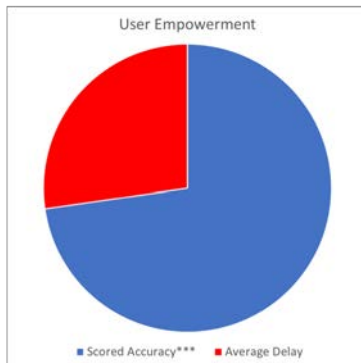Relative Percentage of Variance Explained

- Accuracy was 2x more powerful than delay in predicting User Experience.

- Accuracy significantly correlated to all post-call response items (p<.001).

- Delay **did not** significantly correlate to any post call responses.

- Graphs show the predictive power of scored accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)
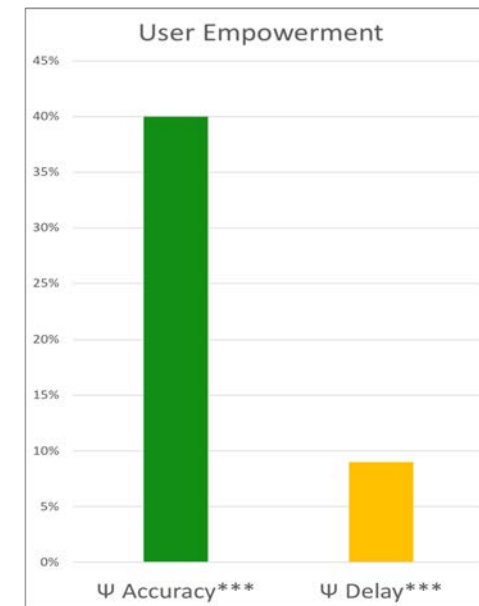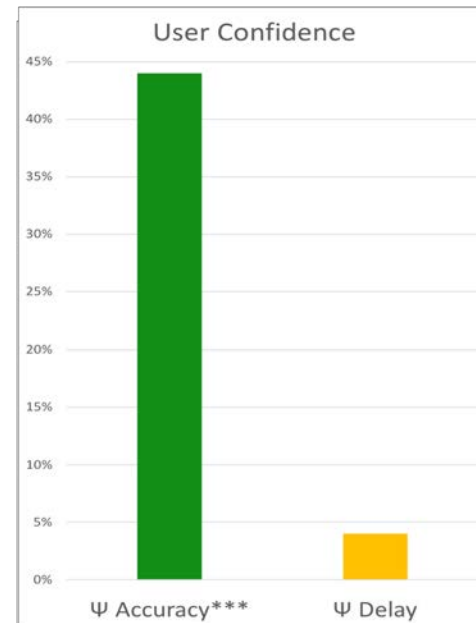
16

Relative Impact on User Confidence*
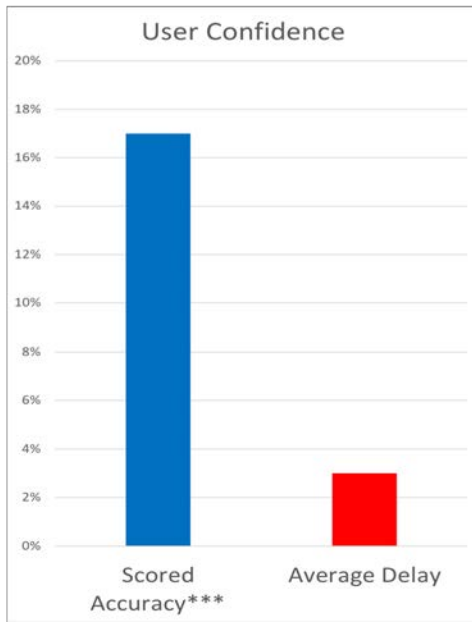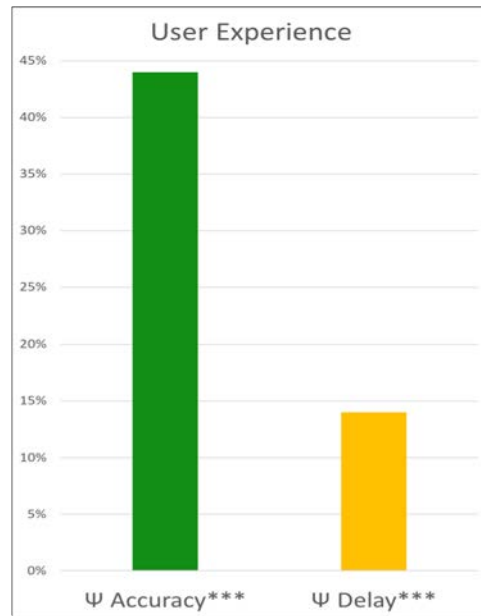

User Confidence

Relative Percentage of Variance Explained
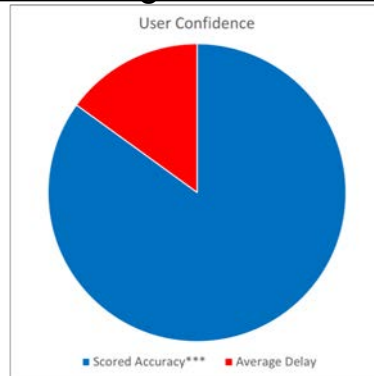
- <u>Accuracy was more than 5x more powerful than delay</u> in predicting User Confidence.

- Accuracy significantly correlated to all post-call response items (p<.001).

- Delay **did not** significantly correlate to any post-call responses.

- Graphs show the predictive power of scored accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)

17

Relative Impact on User Empowerment*
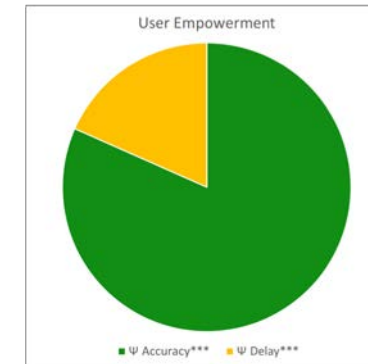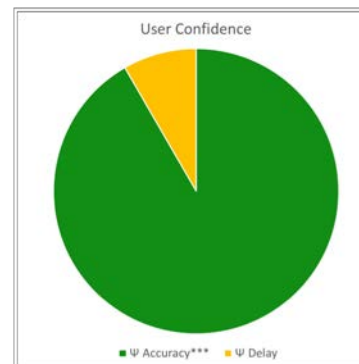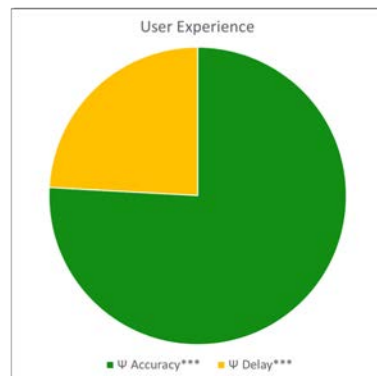

Relative Percentage of Variance Explained

- Accuracy was more than 2x more powerful than delay in predicting User Empowerment.

- Accuracy significantly correlated to all post-call response items ($p<.001$).

- Delay **did not** significantly correlate to any post-call responses.

- Graphs show the predictive power of scored accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)

See what they say

Relative Impact of Perceived Accuracy and Delay*
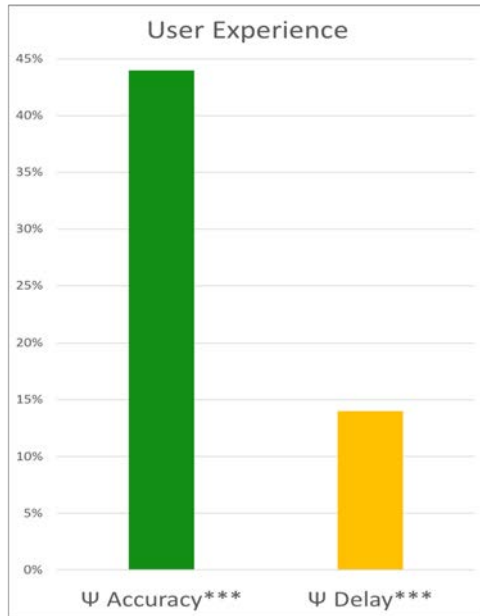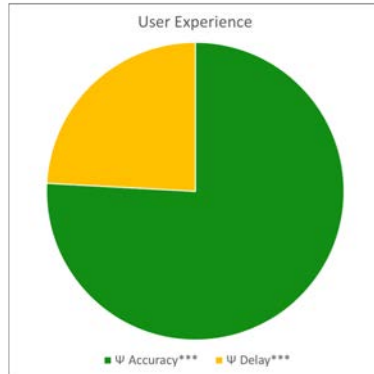

Relative Percentage of Variance Explained

*Comparison of Standardized
Regression Coefficients
(Predictive Power)
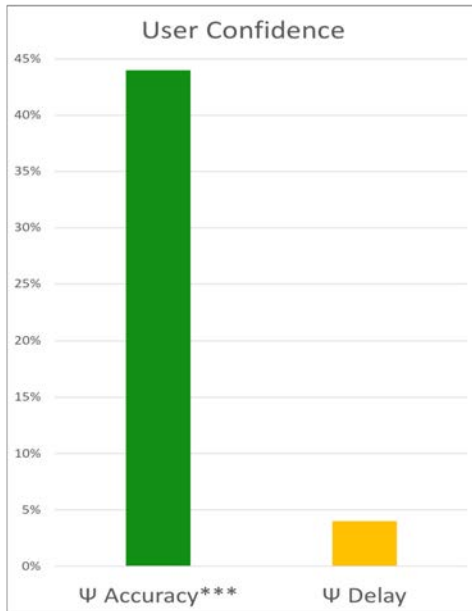
19

Relative Impact on User Experience*

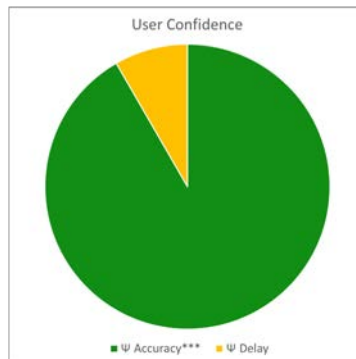Relative Percentage of Variance Explained

- Accuracy was 3x more powerful than delay in predicting User Experience.

- Accuracy and delay were both significantly correlated to post-call response items ($p<.001$).

- Users' perceptions are overwhelmingly based on their experiences of accuracy, over delay.

- Scored accuracy will vary based on tool used, but this correlation is based on what volunteers perceived (users' own assessments).

- Graphs show the predictive power of perceived accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)
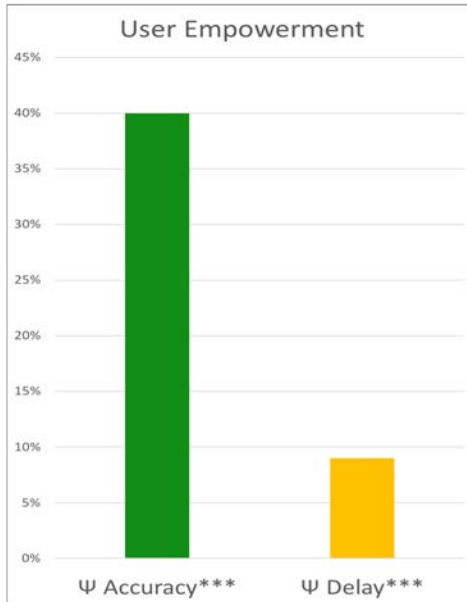
Relative Impact on User Confidence*
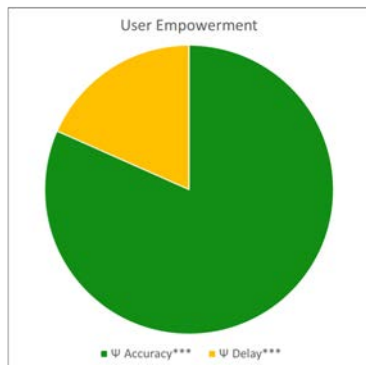

Relative Percentage of Variance Explained

- Accuracy was 10x more powerful than delay in predicting User Confidence.

- Accuracy was significantly correlated to post-call response items ($p<.001$). Delay was not statistically significant.

- Users' perceptions are overwhelmingly based on their experiences of accuracy, over delay.

- Scored accuracy will vary based on tool used, but this correlation is based on what volunteers perceived (users' own assessments).

- Graphs show the predictive power of perceived accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)

Relative Impact on User Empowerment*

User Empowerment

Relative Percentage of Variance Explained

User Empowerment

- **Accuracy was 4x more powerful than delay** in predicting User Empowerment.

- Accuracy and delay were both significantly correlated to post-call response items ($p<.001$).

- Users' perceptions are overwhelmingly based on their experiences of accuracy, over delay.

- Scored accuracy will vary based on tool used, but this correlation is based on what volunteers perceived (users' own assessments).

- Graphs show the predictive power of perceived accuracy and delay on volunteers responses to post-call items.

*Comparison of Standardized Regression Coefficients (Predictive Power)

See what they say®

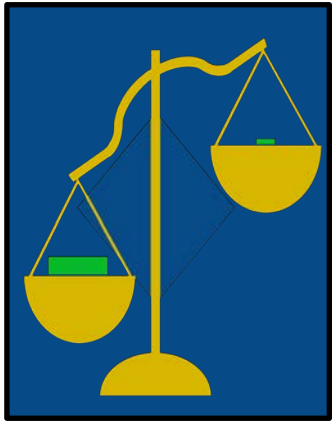# Accuracy Impacts the Perception of Delay
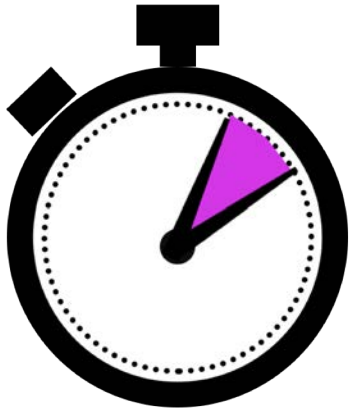
Objective Delay

Perceived Delay

- Some users are more sensitive to delay than others.
- Users frequently perceived accurate captions with greater delay as equal to or faster than inaccurate captions with less delay. However, delay did not seem to impact users' perceptions of accuracy. This may be because inaccuracies tax cognitive load and working memory, impacting the experience of delay and making tasks more strenuous.
- Further Study is Required.

Impacts on IP CTS Quality
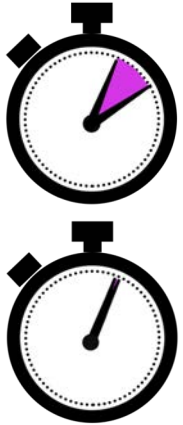
1. **Accuracy Influences Users' Perceptions of Delay**
   - Our research indicates that accuracy influences users' perceptions of delay and therefore poses an independent 'first hurdle' before considering the dependent variable of delay.

2. **Scored Accuracy Significantly Predicts Usefulness (Functionality) of Captions, Delay Does Not**
   - Our research shows that weighing accuracy over speed would better represent usefulness of captions and ensure greater functional equivalence for users.

Thank You