**Before the**

**Federal Communications Commission**

**Washington, D.C. 20554**

In the Matter of                                                          )

Misuse of Internet Protocol (IP) Captioned               )           CG Docket No. 13-24

Telephone Service                                                      )

                                                                                    )

Telecommunications Relay Services and                   )           CG Docket No. 03-123

Speech-to-Speech Services for Individuals                )

with Hearing and Speech Disabilities                         )

**MACHINEGENIUS, INC. RESPONSE TO NOTICE OF INQUIRY**

MachineGenius shares the Commission’s interest in establishing well-defined measures of IP CTS performance which enable consumers to make informed decisions in the selection of IP CTS providers, and allow the IP CTS program to benefit from technological innovation.

We respond herein to the Commission’s Notice of Inquiry issued June 2018, commenting on specific issues raised by the Commission, and offering recommendations for measuring IP CTS performance.

**Summary**

MachineGenius agrees with the Commission on the three primary goals of the IP CTS, namely functional equivalence, leveraging new technology, and improving efficiency, and suggest that a fourth primary goal of customer choice be adopted.

KPIs should be adopted that support the achievement of these goals; these KPIs should reflect the totality of the IP CTS user experience, and as such should include not only measures of technical performance, but also user feedback and tester assessments. The Commission should develop standardized methods of gathering such feedback and assessments.

KPIs should be kept simple, actionable, and easily understood, for the purpose of enabling consumers to make informed choices when selecting an IP CTS provider, and for the purpose of overall IP CTS program evaluation.

The use of new and existing technologies, both specifically related to IP CTS and as alternatives to IP CTS, should be proactively fostered by the Commission. The Commission should especially facilitate the adoption of innovative technologies that have potential to both improve functional equivalence and increase program efficiency.

**Primary IP CTS Goals**

MachineGenius substantially agrees with the Commission that the primary goals of the IP CTS program should be, at minimum:

1. to make communications services available to individuals with communications disabilities that are functionally equivalent to communications services used by individuals without such disabilities;
2. to keep up with technological changes and advances in the telecommunications industry; and
3. improve the efficiency of IP CTS, and reduce the incidence of waste, fraud, and abuse.

We suggest clarifying (1) to reflect that the IP CTS program seeks to provide functionally equivalent *telephonic* communications, and not communications services more generally. This may be an important distinction in determining the extent to which the Commission encourages the usage of alternative technologies such as assistive devices and text- or video-based direct communication.

The Commission should broaden (2) to include changes and advances in areas beyond the telecommunications industry, if such can help make telephony services available to individuals with communications disabilities. For example, the program should take into account changes and advances in audiology, linguistics, artificial intelligence, user behaviors, interface design and usability, and consumer trends, to the extent that these can be applied to advancing IP CTS.

We suggest a fourth primary goal:

1. enabling informed consumer choice in the selection of IP CTS providers.

This goal is already explicit in the Commission’s desire to make performance data available to consumers, and it has been identified as an important goal by the DAC and consumer groups. Consumer choice is virtuous for its own sake, in that it enables consumers to select a service which most completely meets their individual needs. By fostering competition among providers, it also encourages innovation, efficiency, and improvements in functional equivalence – the other primary goals of the program. Consumer choice allows free-market dynamics to reward superior product offerings with higher market share and creates a more robust IP CTS program by mitigating the risk of provider failure.

MachineGenius believes that these goals are appropriate for assessing the IP CTS program and IP CTS provider performance.

**Performance Goals**

*Goal #1: Functional Equivalence*

Functional equivalence is properly the first goal of IP CTS, provided that achieving functional equivalence does not do undue disservice to goal of efficiency. We include this caveat to accommodate the fact that if a method A of IP CTS delivery more nearly achieves functional equivalence than some other method B, but A is prohibitively expensive, then A may not be consistent with program goals overall.

MachineGenius agrees with the Consumer Groups that a reasonable definition of functional equivalence as applied to IP CTS is enablement of “[p]ersons receiving or making relay calls . . . to participate equally in the entire conversation with the other party or parties and . . . experience the same activity, emotional context, purpose, operation, work, service, or role (function) within the call as if the call is between individuals who are not using relay services on any end of the call.”[[1]](#footnote-1)

While MachineGenius agrees that this is a reasonable definition, we note that the goal of functional equivalence is sought “to the extent possible and in the most efficient manner.”[[2]](#footnote-2) It is likely that strict and complete functional equivalence for all possible use-cases is not in principle possible, or is extraordinarily inefficient to achieve, and that what constitutes “practical” functional equivalence is *dependent upon the goals of the end user*. For this and other reasons detailed below, we recommend that at least some IP CTS metrics be derived from user ratings of functional equivalence, and not exclusively from aspects of technical performance.

*Goal #2: Technological Advances*

MachineGenius agrees that the second goal of the IP CTS program should be to ensure that the program utilizes technological changes and advances in the telecommunications industry (and others) to the greatest extent possible in service of achieving functional equivalence. This goal follows naturally from goals (1) and (3), since the goal of technological advances in any domain is to improve performance and efficiency.

It is not clear that the Commission is under mandate to foster adoption and outreach related to off-the-shelf and assistive technologies, since the definition of TRS in Section 225 applies specifically to telephone transmission services[[3]](#footnote-3). Likewise, text- and video- based direct communication alternatives do not specifically fulfill the requirement or Section 225, since these are not telephonic in nature. However, we believe that the FCC is in a privileged position to foster such adoption; these alternatives can and do address the needs of hard-of-hearing users in many circumstances, and it is likely to be in the interests of the Commission to encourage use of these technologies for the purposes of reducing the burden on the TRS fund.

Conversely, the FCC is required per Section 225[[4]](#footnote-4) to encourage use of existing of technology, and not to impair the development of improved technology, that does specifically relate to telephonic communication. These include ASR, hi-definition codecs, broadband and cellular services, VoIP services, and amplified telephones, among others. We recommend that the Commission provide expedited processes for the investigation, trial, integration, and adoption of technologies in this class.

*Goal #3: Provision of Service in the Most Efficient Manner*

It is appropriate that the third program goal be improving the efficiency of the IP CTS program and to reduce the program’s incidence of waste, fraud, and abuse. The key measure of efficiency should be cost per reimbursable minute to deliver IP CTS service, and there should also be measures of how many people needing IP CTS are actually being served by it.

There may be a tradeoff between efficiency and level of functionally equivalent performance, because it may be that it costs more per minute to deliver better performance. MachineGenius believes that this highlights the role of innovative technology: new technologies have the potential to eliminate this tradeoff by delivering improved functional equivalence simultaneous with efficiency gains, and their adoption should be proactively facilitated in service of goal (3).

*Proposed Goal #4: Consumer Choice*

As noted above, consumer choice not only benefits end-users but encourages competition crucial to the success of the IP CTS program, and therefore merits inclusion as a primary goal of the program.

**Performance Measures**

Performance measures in the form of KPIs (Key Performance Indicators) should be adopted for assessing progress toward IP CTS program goals. Use of KPIs is an approach with demonstrated efficacy across multiple industries. KPIs may be different for program-level vs provider-level performance, but in both cases should be: associated with a specific objective, relatively few in number, actionable, discriminative, and easily understood. For purposes of measuring functional equivalence, KPIs should reflect bottom-up technical system performance as well as top-down user experience ratings and customer satisfaction as gathered from user feedback. A proper choice of KPIs will suffice for both evaluation of provider performance and providing data to consumers to make an informed choice among providers.

Testing procedures should be performed primarily by an independent 3rd party or parties, though certain tests and data gathering may appropriately be performed by providers themselves. The TRS fund administrator should oversee the scheduled aggregation and dissemination of this data in the form of KPIs. Metrics should be published on a consumer-friendly website maintained by the FCC or by the independent 3rd party evaluator(s).

Because of the importance of user feedback in evaluating service quality, a standardized process by which IP CTS users can rate the quality, performance, and experience of IP CTS calls should be implemented. This can take the form of in-application / on-device feedback affordances, or periodic phone, printed, or electronic surveys by an independent 3rd party chosen by the Commission.

**Functional Equivalence**

The Commission seeks comment on the use of the following metrics to measure IP CTS service quality: (1) transcription accuracy; (2) transcription synchronicity; (3) transcription speed; (4) speed of answer; (5) dropped or disconnected calls; (6) service outages; and (7) usage data, including how often these should be tested or gathered, and how information gathered should be reported.

MachineGenius suggests that these seven metrics do not without modification appropriately measure IP CTS functional equivalence. We recommend instead, within the above-mentioned framework of KPIs, a mix of technical performance measures and user and/or tester ratings assessments that more accurately represent the level of functional equivalence of an IP CTS service. We note also that “service quality” encompasses the entirety of the user experience, and not just the functional equivalence of a product offering. Usability and interface design, feature set, customer service, call quality, service reliability, mobility, and likely other factors contribute to a consumer’s overall satisfaction with a service.

The importance of user and tester ratings cannot be overstated. The case can be made that a user’s answer to “How satisfied are you that the solution has offered you functional equivalence?” is of more value than any technical measure of performance, and is possibly the only thing that matters. Furthermore, user feedback can capture hard-to-quantify aspects of performance, can help discover unknown factors impacting performance, and offers a chance to correlate underlying technical measures with top-down holistic assessments. Modern product and service companies consider user feedback to be of the highest value, and the IP CTS program should follow this lead and incorporate measures of this type among program KPIs.

We suggest developing a Provider Performance Scorecard which includes the following KPIs intended to measure overall service quality and functional equivalence:

1. Accuracy
2. Understandability
3. Delay
4. Reliability
5. Usability
6. Customer Service
7. Overall User Experience

These are discussed below.

*Accuracy.*

Caption accuracy is a critical component of functional equivalence, and MachineGenius agrees that accuracy should measured on a regular basis and included as an IP CTS KPI.

The Commission asks whether accuracy guidelines developed for closed captioning are appropriate for IP CTS and whether guidelines, if any, for CART, court reporting, or related activities can be adapted for use with IP CTS. We believe that while the spirit of closed captioning accuracy rules applies to IP CTS, the rules as written are overly complex and not ideally suited to IP CTS, and that borrowing rules from CART or court reporting likewise presents unnecessary complications.

For the purposes of IP CTS, a straightforward measure of major-Word Accuracy Rate (WAR) is an appropriate proxy for the intuitive notion of “accuracy” and will best serve both the Commission and IP CTS users. We submit that this KPI be named simply “Accuracy”, and agree with the definition proposed by current IP CTS providers[[5]](#footnote-5):

*Accuracy for IP CTS is defined as 100% minus the Major Word Error Rate on the final displayed captions, where the Major Word Error Rate is the number of word substitutions, deletions, and insertions that significantly alter, obscure or reverse the meaning of the original speech divided by the total number of words in the original speech.   It is expressed as a percentage between 0 and 100%.*

MachineGenius also agrees with the providers that a number of presentation-level details should not impact the Accuracy measure[[6]](#footnote-6):

* Uppercase/Lowercase and punctuation are not considered in the accuracy calculations.
* Contractions and expanded phrases are both valid (“that’s” and “that is” are considered the same).
* Abbreviations that have spaces or periods between the letters are considered valid (“FCC”, “F C C”, and “F.C.C.” are all considered the same.
* Numbers may be spelled out or numeric (“400”, “four hundred”, and “4 hundred” are all considered the same).
* Times may be represented as words or numbers (“8:30”, “8 30”, “830”, and “eight thirty” are all considered the same).
* Hyphenated words, non‐hyphenated words, and words separated by underscores are all valid (“thank you”, “thank‐you”, and “thank\_you” are all considered the same).
* Universal Resource Locators (URL) that contain extra spaces or spell the words “slash” or “dot” are valid (“fcc.gov/smartdevice”, “fcc dot gov slash smart device”, and “fcc . gov / smartdevice” are all considered valid).
* Disfluencies (“ah”, “um”, “hmm”) may be omitted, but are not counted as errors if included.
* Address abbreviations are valid. For example, “South 16th East” and “S. 16 E.” are considered the same.

*Understandability***.**

Although the presentation-level details above are not included in the calculation of Accuracy, they may have a substantial impact on the user experience, and we suggest an Understandability KPI based on the subjective ratings of IP CTS users and 3rd-party testers. The Understandability KPI should be based on top-down user/tester ratings, rather than an algorithmic, bottom-up metric based on some weighting of measures of the listed items above, or any other such list. The user-rating approach is favorable in that it can incorporate elements of presentation not specifically enumerated, like punctuation, timing information, representation of silence, prosodic information, and represents the overall integration of all presentation-related elements and their potential interaction. This would be difficult to achieve with a bottom-up approach.

For the purposes of KPIs, Accuracy and Understandability measurements should be performed across a variety of call types reflective of the distribution of call types in the wild and can optionally be broken down on a per-call-type basis, but the methodology of measurement should not be adjusted for type of call.

We believe that measures of accuracy beyond that described above will have diminishing returns and may be subject to unproductive debate. We do not believe that “completeness”, as suggested by MITRE[[7]](#footnote-7), adds any value since incomplete captioning will already be captured in the Accuracy score (in the form of deletions). Using semantic error rate as an assessment of accuracy is likely to introduce methodological complexity disproportionate to any incremental improvement versus Accuracy as defined above.

It is not necessary to develop a methodology which measures Accuracy in real-time on live calls in order to comply with the prohibition on call content retention. Accuracy can be evaluated by anonymous callers and callees, employed by independent third parties, making and recording both scripted calls and open-ended calls in the wild. The recording of audio and caption data necessary for Accuracy and Understandability metrics can be performed by the tester without violating retention rules.

Testing, gathering of data, and reporting for Accuracy and Understandability should occur quarterly. Monthly measurements, as have been proposed elsewhere, will be overly burdensome and are unlikely to provide significant incremental benefit, since levels of accuracy and understandability are unlikely to experience large deltas over short time periods. For example, if a low Accuracy score is found, it is unlikely that any ameliorative action can be executed within a single month.

Intra-quarter testing can also be done on an ad hoc basis upon new product/feature release, upon report of substantive change, or by request of providers.

Accuracy and Understandability KPIs should apply equally to captions generated by ASR, CAs, or by any other method or combination of methods.

*Delay*

Synchronicity and Transcription Speed are two (closely related) measures that are critical to functional equivalence and should be measured by the Commission. Because transcription speed will necessarily be reflected in measures of synchronicity, we suggest combining these two into a single metric called “Delay” that measures the average duration between when a given word is uttered and when that word is accurately or finally transcribed (whichever comes last).

Like Accuracy, Delay can be measured based on audio and caption recordings captured by third-party testers. Both parties on the call can be testers, if necessary, and it is not necessary to measure delay in real-time on live calls.

Delay does not need be measured on every call; appropriate sampling should be sufficient. Like accuracy, Delay is unlikely to vary over short periods of time. Factoring of causes for latency is an unnecessary complexification.

*Reliability*

We suggest that metrics pertaining to speed of answer, dropped or disconnected calls, and service outages be combined into a Reliability KPI for consumer use, though they should be tracked independently for FCC internal purposes. The Reliability KPI should be calculated as successful calls divided by total calls, where a successful call must be answered within 15 seconds and must not be dropped or disconnected. Estimated call volumes based on historical data should be used for calculating unsuccessful calls due to outages. These metrics should be gathered by providers and reflect every call, not a sample of calls.

*Usability, Customer Service, and Overall User Experience*

Like Understandability, these KPIs should be based on user and/or tester ratings, and not measures of technical performance. They should be gathered via standardized mechanisms prescribed by the Commission. Although these KPIs are subjective on a per-user basis, in the aggregate they provide a quantifiable measure, allowing crucial visibility into important aspects of service quality, and permitting consumers to make informed choices when selecting an IP CTS provider.

*Usage Data*

The Commission should collect from every provider not only minutes used, but also call count, unique user count, and the distribution of call count and minutes used per user. This can be achieved while keeping user PII confidential and without reference to TRS URD. Anonymous demographic data such as age, gender, severity of hearing loss, and self-reported anticipated uses of IP CTS (business, medical, personal, etc.) should be collected by providers and periodically updated to the extent possible. The Commission should consider making the collection of such data required at user signup.

**Technological Advances**

MachineGenius agrees that new (and existing) technologies such as text, video, and RTT can provide effective alternatives or supplements to IP CTS in certain situations. As above, we believe that these technologies naturally fall into two classes: those which directly bear on the functional equivalence of telephonic communication (where a user can make and receive calls to and from PSTN/cellular networks via the technology in question), and those which represent alternatives to telephony appropriate to certain circumstances. The Commission is required to foster the use of technologies that fall into the former class; HD / wide band audio codecs are an example, as is ASR. However, even though not required to address alternative, non-telephonic technologies, MachineGenius believes it is in the interests of both the hard-of-hearing community and the TRS fund for the Commission to make assessments of the efficacy of such technologies, for the purposes of outreach to benefit consumers who may not be aware of these technologies, and the purpose of maintaining the integrity of the TRS – by ensuring that IP CTS is only utilized where it the solution that best benefits the consumer.

There exists a minimum level of performance (or alternatively, user experience) below which off-the-shelf, assistive, and alternative technologies should not be encouraged. We suggest that at minimum, performance should be such that a) the safety and well-being of users is not compromised, b) users do not bear undue cost in the form of monetary expense, user effort, time, or switching costs (it should be easy to “try out” alternative solutions, and not have vendor or technology lock-in), and c) user experience can reasonably be expected to be at least as good as current mainstream solutions.

**Program Efficiency**

The Commission seeks comment on metrics appropriate to measuring the efficiency of the IP CTS program. We suggest these are the appropriate KPIs: 1) reimbursement cost-per-minute, 2) estimated % of market served, 3) estimated % waste, and 4) the market-share-weighted average of the per-provider Overall User Experience KPI.

1. Letter from Tamar Finn, Counsel to Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), to Marlene H. Dortch, Secretary, FCC, CG Docket Nos. 03-123 and 10-51, Attach. at 1 (filed Apr. 12, 2011) (“Consumer Groups’ TRS Policy Statement – Functional Equivalency of Telecommunications Relay Services: Meeting the Mandate of the Americans with Disabilities Act”). [↑](#footnote-ref-1)
2. 47 U.S.C. § 225(a)(3), (b)(1). [↑](#footnote-ref-2)
3. *Ibid.* [↑](#footnote-ref-3)
4. 47 U.S.C. § 225(d)(2). [↑](#footnote-ref-4)
5. IP CTS Quality Metrics: Provider Recommendations; August 21, 2018. [↑](#footnote-ref-5)
6. *Ibid.* [↑](#footnote-ref-6)
7. See MITRE Corporation, Internet Protocol Caption Telephone Service (IP CTS) Devices: Summary of Phase 1 Activities (2017), CG Docket Nos. 03-123 and 13-24 at 4-6 (posted Apr. 11, 2018), https://ecfsapi.fcc.gov/file/10411287298464/MITRE%20Corporation%20Summary%20of%20Phase%201.pdf (MITRE Phase 1 Summary) (measuring completeness and accuracy separately). [↑](#footnote-ref-7)