# Readability of Punctuation in Automatic Subtitles

Promiti Datta[2], Pablo Jakubowicz[1], Christian Vogler[1] and Raja Kushalnagar[1]

[1] Gallaudet University, Washington, DC 20002, USA
[2] University of Toronto, Toronto, Ontario, Canada

**Abstract.** Automatic subtitles are widely used for subtitling television and online videos. Some include punctuation while others do not. Our study with 21 participants watching subtitled videos found that viewers reported that punctuation improves the "readability" experience for deaf, hard of hearing, and hearing viewers, regardless of whether it was generated via ASR or humans. Given that automatic subtitles have become widely integrated into online video and television programs, and that nearly 20% of television viewers in US or UK use subtitles, there is evidence that supports punctuation in subtitles has the potential to improve the viewing experience for a significant percentage of the all television viewers, including people who are deaf, hard of hearing, and hearing.

**Keywords:** Subtitles, Subtitles, Deaf, Hard of Hearing.

## 1 Introduction

Subtitles allow translation of auditory information into a visual representation on the screen. Subtitles give all viewers, including those who are deaf or hard of hearing a visual medium to follow video content that includes an auditory track. Improving the availability and accuracy of subtitles offers benefits for everyone, regardless of whether they are deaf, hard of hearing or hearing or not. In fact, nearly 20% of the population in the US or UK use subtitles; and 80% of them are hearing [3,4]. Many hearing viewers who watch subtitles do so because they are learning English as a second language or watching TV in noisy settings such as pubs. Before the widespread adoption of Automatic Speech Recognition (ASR), subtitles for television, education, or courtroom reporting were generated by human-powered subtitling services such as stenography or re-speaking, that usually generated punctuated subtitles [2,6,7]. So, the issue of evaluating punctuation versus unpunctuated subtitles was not considered until the advent of ASR.

Live subtitling is challenging, as the text needs to be produced immediately, with almost no time for reaction and correction. The accuracy of ASR services with low latency has vastly increased the amount of television programming that can be subtitled. However, some ASR services include punctuation while others do not. While it seems intuitively true that subtitles will be harder to read without punctuation, this has not been widely investigated, because human-generated subtitles are usually punctuated. Our study investigates how punctuation in subtitles is related to ease of reading and contributes to the overall "readability" experience. It compares viewer experiences for both human and ASR generated punctuated and unpunctuated subtitles.

## 2 Related Work

ASR is being integrated into television and video streaming services. For example, YouTube offers 'automatic subtitles' using its ASR services. Other streaming services use Google's 'Cloud Speech-to-Text'[1], Microsoft's 'Speech Services'[2], or Amazon's 'Amazon Transcribe'[3], and video players are integrating ASR in their options. While ASR is fast, its performance in transcribing and punctuating live speech has been less accurate than transcribing pre-recorded speech, as the machine has less time to make decisions on what has been said and is unable to take the words that follow an utterance into account. However, as automatic speech recognition services have become more accurate and complex, these services have begun to incorporate reliable automatic punctuation into their transcriptions, through a combination of lexical and prosodic features, such as pause length in speech. In live stenography for television, the stenographers utilize the same training as a court stenographer [2], with. For live television subtitling, speakers tend to speak with less structure and more variance than in court, so subtitling quality is usually inferior to court reporting. The subtitling quality is affected by the delay of a human stenographer's or re-speaker's response in listening and transcribing live speech, and usually has a higher error rate due to transcribing under pressure [1,5].

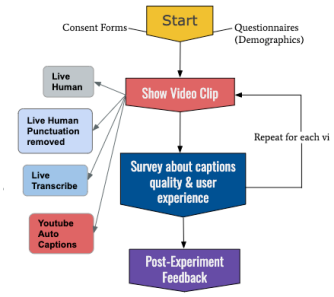
[1] https://cloud.google.com/speech-to-text/
[2] https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/
[3] https://aws.amazon.com/transcribe/

# 3 Methods

## 3.1 Video

For the study, we gathered four videos with different subtitle generation and formatting: 1) punctuated subtitles generated by re-speakers, 2) unpunctuated subtitles generated by re-speakers, 3) punctuated subtitles generated by Google Live Transcribe[4], and 4) YouTube Automatic Subtitles[5]. All video clips were taken from live television broadcasts. Each video was trimmed into a two to four-minute clip that contained one segment of the television show. For the purpose of this experiment, the clips were categorized into news segments, and talk show segments.

*Table 1: Methodology*

| Steps | Stimuli Categories | |
|---|---|---|
| | NO Punctuation | Punctuation |
|  | Automatic Speech Recognition |  **YouTube** "a good feeling we" |  **Live Transcribe** (Google's app) "a good feeling." |
| | Human |  **Live Human 0** (no punctuation) "GOOD FEELING >>WE WERE ALL JUST" |  **Live Human** "GOOD FEELING." |

## 3.2 Types of Subtitling

**Live Human Generated Subtitles with and without punctuation.** Television networks hire professional stenographers or re-speakers to generate live subtitles as a show is broadcasted. The subtitles are comprehensive and contain very few grammatical mistakes but may use paraphrasing and do not follow the exact words of the speaker. The subtitles were downloaded from recorded videos using CCExtractor. For these subtitles, the re-speakers have explicitly added punctuation. For this experiment, one version of the subtitles contained all the original punctuation, and one had all commas, periods, exclamation marks, and question marks removed.

[4] https://www.android.com/accessibility/live-transcribe
[5] https://support.google.com/youtube/answer/6373554?hl=en

**Auto-generated Subtitles with** and **without punctuation.**
The auto-generated subtitles were made using either Google's YouTube auto-subtitling service, or Google's Live Transcribe app. Both had similar word accuracies for each of the videos, however Live Transcribe provided automatic punctuation and YouTube subtitles did not.

*Table 2: Subtitle Type*

| | Human | ASR |
|---|---|---|
| Punctuation | Live Human Subtitles | Live Transcribe |
| No Punctuation | Live Human Subtitles, punctuation removed | YouTube Auto-generated Subtitles |

## 4 User Evaluation

### 4.1 Participants

The experiment involved 21 participants, out of which 12 identified themselves as men and 9 as women. Nineteen participants identified as Deaf, and two as Hard of Hearing. Participants self-reported being mostly fluent in both ASL and English and were mostly ages 18 to 40.

### 4.2 Experiment Design

The participants completed a consent form and demographic survey. Next they viewed four video clips 2-4 minutes long, in random order. After each video, the participants completed a short survey including quantitative questions on a Likert scale, and answered qualitative questions. After they watched all videos, the participants were invited to provide feedback doing an overall comparison of all the video subtitles.

*Table 3: Counterbalancing*

| G1 | G2 | G3 | G4 |
|---|---|---|---|
| LH | LT | $LH_0$ | YT |
| LT | $LH_0$ | YT | LH |
| $LH_0$ | YT | LH | LT |
| YT | LH | LT | $LH_0$ |

Each of the subtitle types - live human (LH), live human with no punctuation ($LH_0$), YouTube (YT), and Live Transcribe (LT), were counterbalanced using a Latin Square design with four groups. In each group, the type of video (news or talk show) alternated for each video.

# 5 Study Results

The videos were chosen to be easy to follow along, so that the viewers could focus on punctuation quality and readability, and not be distracted by too many incorrect words or difficult content. After completing the video section of the experiment, participants were asked to give short answers describing their experience reading punctuated subtitles vs. non-punctuated subtitles, as well as how much they were able to tolerate punctuation errors. There were common themes across most answers, as 16 out of 21 participants reported that they prefer some level of punctuation in subtitles over no punctuation.

## 5.1 Ability to follow along

Participants reported their ability to follow along with each type of subtitle as shown below. Most people were able to follow along with all four subtitle types, reporting scores of mostly fours and fives, with a bit more variability for the non-punctuated subtitles (YT and $LH_0$), as shown below, did not have a big impact.

## 5.2 Subtitle Readability by Format

Punctuated subtitles (LT and LH) were easier to read than non-punctuated subtitles ($LH_0$ and YT), which had more variability, but overall harder to read. Even with the same exact words, getting rid of punctuation had a noticeable negative impact on readability.

## 5.3 Comparison

Punctuated subtitles (Live Human and Live Transcribe) had a positive impact on readability compared to non-punctuated subtitles (LH - No Punctuation and YouTube). Many people did not feel the lack of punctuation had a great impact on readability.

## 5.4 Punctuation in Human Generated Subtitles

In a 2-sample t-test against overall readability scores for Live Human written subtitles versus the same subtitles with the punctuation removed, there was not a significant difference between scores.

### 5.5 Punctuation in Auto-Generated Subtitles

The 2-sample t-test between readability scores for the punctuated Live Transcribe versus non-punctuated YouTube subtitles, there was a significant preference towards the punctuated subtitles.

Average Ratings by Caption Type

| Question | LH | $LH_0$ | YT | LT |
|---|---|---|---|---|
| Ability to Follow Along with Captions | 4.6 | 3.6 | 4.2 | 4.4 |
| Comprehension of Video | 4.5 | 3.9 | 3.9 | 4.2 |
| Word Accuracy | 4.3 | 3.9 | 3.7 | 4.1 |
| Readability and Grammar | 4.1 | 3.3 | 2.7 | 4.0 |

### 5.6 Human Generated Subtitles vs. Auto-Generated Subtitles

There was a stronger preference toward human generated subtitles when human and auto generated subtitles either both had or punctuation or both did not have punctuation. A comparison showed significant preference for the punctuated auto-generated subtitles. Overall, subtitles with punctuation scored significantly higher than non-punctuated subtitles.

The figure above highlights the success of the punctuated Live Transcribe (LT) subtitles over the non-punctuated human subtitles ($LH_0$). These conclusions are further reinforced by a follow up question asked of all participants: “Would you prefer punctuated subtitles with some punctuation errors over non-punctuated subtitles?”, 83% said yes.

Common feedback with non-punctuated subtitles was that participants had trouble understanding grammar and sentence structure, run-on sentences, and identifying who was talking. Four people said non-punctuated subtitles were too hard to read.

Participants’ comments generally indicated that they spent less effort on reading subtitles with punctuation. “When I read the non-punctuated subtitles, they look like run-on sentences and I have a hard

time trying to figure out when they stop talking," "[Punctuated Subtitles] also help me separate concepts, sentences, paragraphs and so on. It makes everything much more transparent". "For non-punctuated subtitles, it made me lose motivation to understand everything because I lost track," "[Non-Punctuated Subtitles] wear my eyes out when I keep reading and notice there is no period. It affects my writing and reading skills".

## 6 Conclusion

Many DHH users prefer perfect subtitling with punctuation over no punctuation. In some cases, proper punctuation had a greater impact on readability than higher word accuracy. In the survey results and feedback, they mentioned that subtitles with punctuation errors are harder to read due to run-on sentences, not being able to tell who is speaking, and difficult reading complex sentence structures. Punctuation plays a very important role in conveying intended meaning to the language. Errors in punctuation or even wrong placement can change the meaning of the sentence completely and sometimes convert to confusion. Viewers reported they found it much easier to follow subtitles with punctuation.

## 7 Acknowledgements

We thank the National Science Foundation, grant #1757836 (REU AICT) and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR #90DPCP0002). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this paper do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.

## References


1. Marta Arumi-Ribas and Pablo Romero-Fresco. 2008. A Practical Proposal for the Training of Respeakers. JoSTrans: The Journal of Specialised Translation 10: 106–127.
2. Greg Downey. 2006. Constructing "Computer-Compatible" Stenographers: The Transition to Real-time Transcription in Courtroom Reporting. Technology and Culture 47, 1: 1–26. https://doi.org/10.1353/tech.2006.0068
3. A B Jordan, A Albright, A Branner, and J Sullivan. 2003. The state of closed captioning services in the United States. Washington, DC. Retrieved from https://dcmp.org/learn/static-assets/nadh136.pdf
4. United Kingdom Ofcom. 2006. Television Access Services: Review of the Code and Guidance.
5. Pablo Romero-Fresco and Juan Martinez. 2011. Accuracy Rate in Live Subtitling – the NER Model. Retrieved from https://roehampton.openrepository.com/roehampton/bitstream/10142/141892/1/NER-English.pdf
6. M. S. Stinson, L. B. Elliot, and R. R. Kelly. 2008. Deaf and Hard-of-Hearing Students' Memory of Lectures with Speech-to-Text and Interpreting/Note Taking Services. The Journal of Special Education 43, 1: 52–64. https://doi.org/10.1177/0022466907313453
7. M Wald. 2005. Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality. In Frontiers in Education, 2005. FIE '05. Proceedings 35th Annual Conference, 22–25. https://doi.org/10.1109/FIE.2005.1612286