

Speech Recognition Software for Audiovisual Translation

Mara Campbell, CT
COO of True Subtitles LLC

Definitions

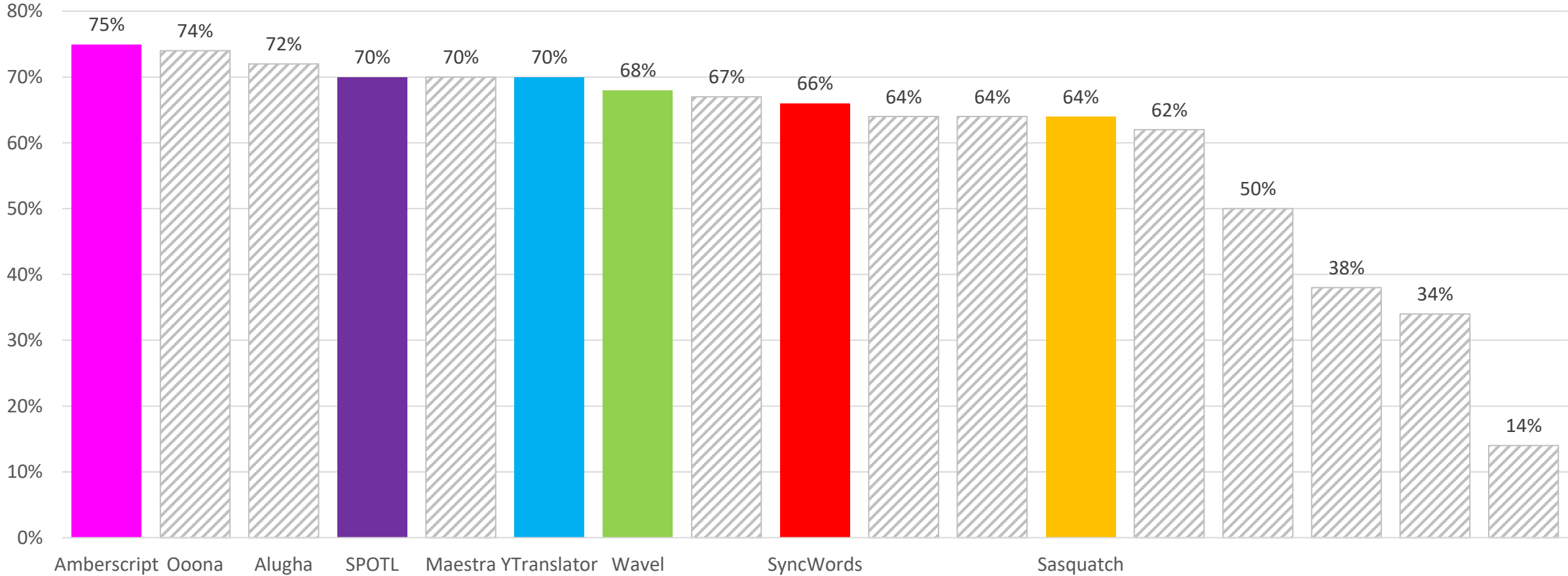
- ASR = automatic speech recognition = speech-to-text (STT).
- Speech processing engines: Google, IBM Watson, Microsoft, Amazon, Speechmatics, Apptek.
- Product: a timed transcription split into subtitles/captions in formats like SRT, STL, etc.
- Benefits: Cut costs and improve turnaround times.

Stage 1

- Sample video: 10-minute clip of *Brooklyn Nine-Nine*.
- Processed through 22 ASRs
- Compared the accuracy of the transcriptions to the script.



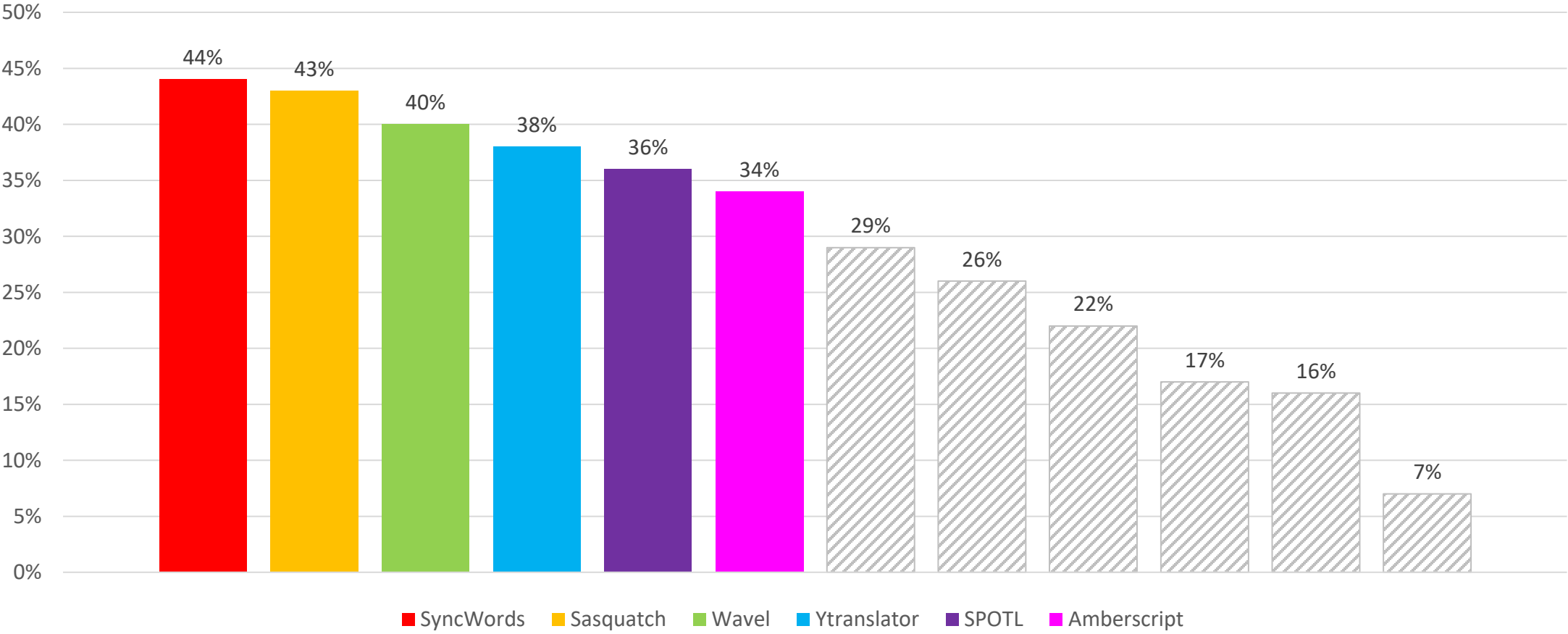
Accuracy



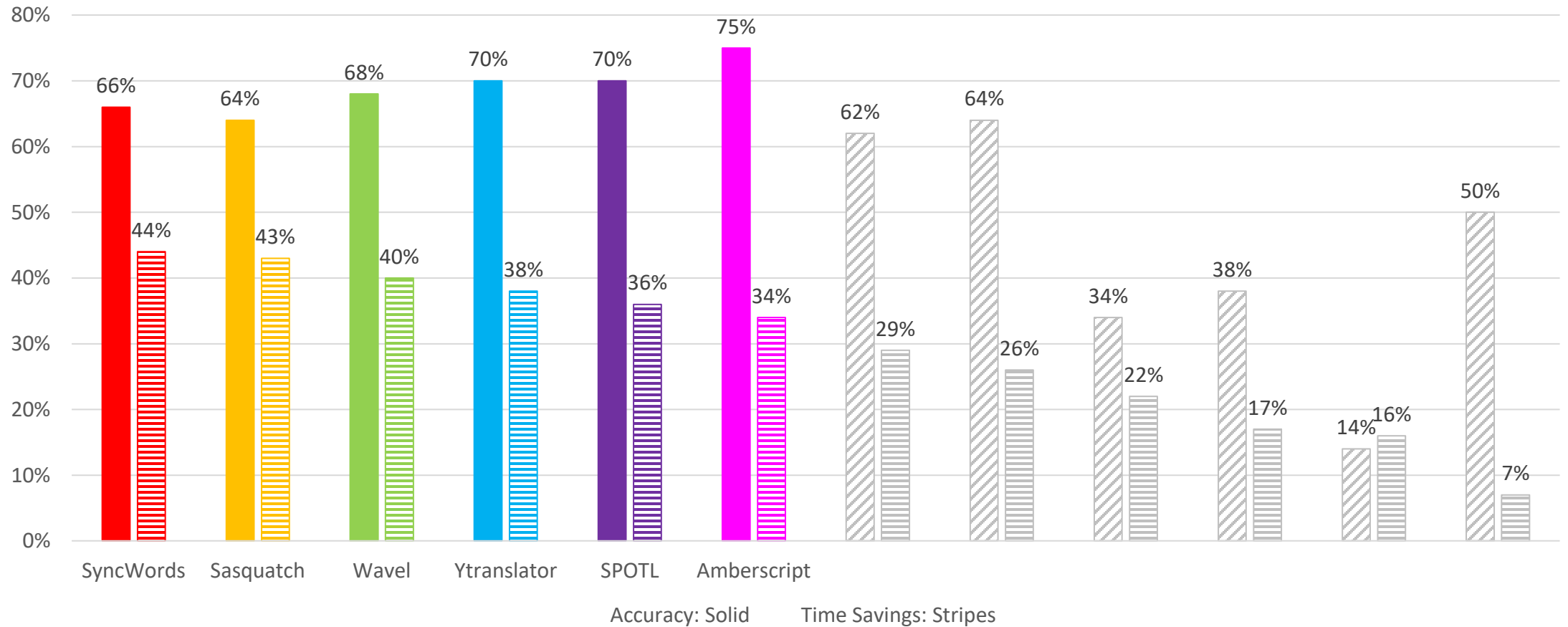
Stage 2

- Timed the operator doing the origination.
- Timed the operator doing the conforming of the ASR output files.
- Calculated the time saved on conforming compared to origination.

Time savings



Accuracy vs. Time Savings



Findings 1: Factors affecting time-savings

1. Punctuation and capitalization.
2. Line and subtitle breaks.
3. Timing.
4. Transcription accuracy.

Findings 2: Type of material

- **Best:** single speaker; scripted/rehearsed material.
- **Worst:** multi-speaker; unscripted material; unclear audio.

Integration

- Ooona and SubtitleNext.
- They both work with the same engines, with an accuracy of 70%.
- The video is uploaded to only one platform, thus bound to its security measures and certifications.

Software features

- Text editing tools.
- Timing editing tools.
- Glossary upload.
- Script upload and “alignment”.
- Spec configuration.
- Human post-editing.
- Project management features.
- MT.

Machine Translation: Considerations

Two approaches:

- On the subtitle level.
- On the paragraph/speaker level.

Editing in the ASR platform 1

- Edition is made directly onto the subtitles.
- Spec errors are indicated.

The screenshot displays a video player interface for editing subtitles. The video shows a man in a white shirt and tie. A subtitle is highlighted in yellow: "La otra pregunta sería OK, hablamos del distrito ahora, solamente en esta iglesia,". A tooltip above the subtitle states: "The subtitle has the following problems" and "The subtitle is 5.17 seconds long. Try to keep the subtitle visible for at most 5 seconds." The right sidebar shows a list of subtitles with their start and end times and a "COMPLETED" status. The bottom of the screen features a timeline and playback controls.

BACK PU0322DIG_0067.mp4

LANGUAGE: SPANISH

EXPORT ✓ COMPLETED

00:01:44:01 00:01:49:23
Artificial
o final que se estaba llevando a cabo?

00:01:49:23 00:01:52:13
Sí, sí. No, no, no, no!

00:01:52:13 00:01:57:18
Estuve dando los detalles y poniendo en silencio. Sí, sí. Dígame usted.

00:01:57:18 00:02:02:10
La otra pregunta sería OK, hablamos del distrito ahora, solamente en esta iglesia

00:02:02:10 00:02:04:26
acá en donde estamos.

00:02:04:26 00:02:08:20
Cuántos miembros está reunido en este momento en esta iglesia? Acá?

00:02:08:20 00:02:12:15
No sé. Confía. La iglesia de la Tilly o. Iglesia de la

+ Add Subtitle

00:01:57:18

01:52 01:53 01:54 01:55 01:56 01:57 01:58 01:59 02:00 02:01 02:0

Estuve dando los detalles y poniendo en silencio. Sí, sí. Dígame usted. La otra pregunta sería OK, hablamos del distrito ahora, solamente en e... acá en de

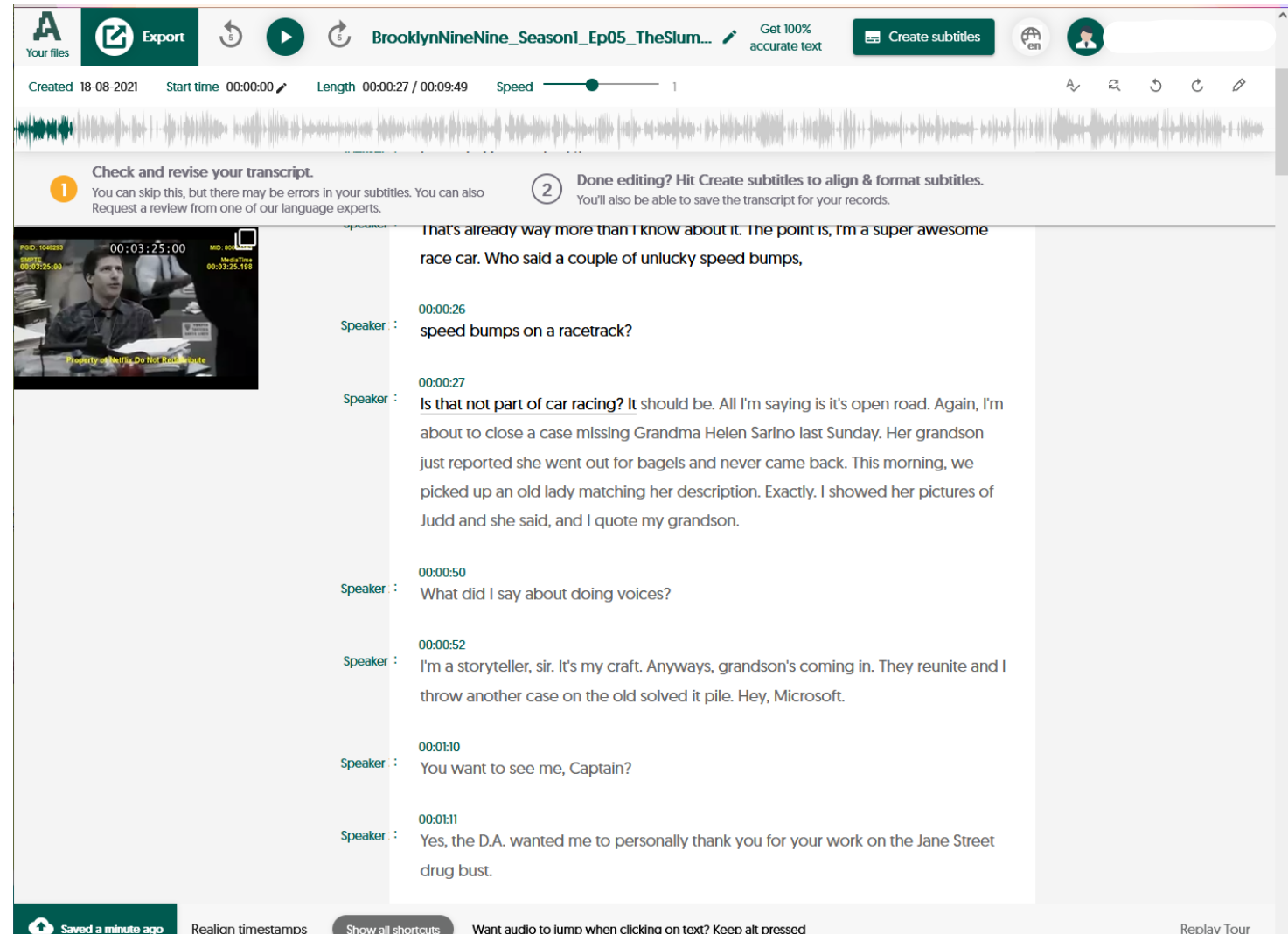
00:01:57:18 TC

SET IN SET OUT

help

Editing in the ASR platform 2

- The text is split into speakers.



The screenshot displays the ASR platform interface for a video titled "BrooklynNineNine_Season1_Ep05_TheStum...". The interface includes a top navigation bar with an "Export" button, a "Get 100% accurate text" badge, and a "Create subtitles" button. Below the navigation bar, there is a progress bar showing the video's length (00:00:27 / 00:09:49) and a speed control slider. A waveform visualization of the audio is visible above the transcript.

The transcript is displayed in a light gray box with a white background, showing the following segments:

- 1** Check and revise your transcript. You can skip this, but there may be errors in your subtitles. You can also Request a review from one of our language experts.
- 2** Done editing? Hit Create subtitles to align & format subtitles. You'll also be able to save the transcript for your records.

The transcript content is as follows:

00:03:25:00
Speaker : That's already way more than I know about it. The point is, I'm a super awesome race car. Who said a couple of unlucky speed bumps,

00:00:26
Speaker : speed bumps on a racetrack?

00:00:27
Speaker : Is that not part of car racing? It should be. All I'm saying is it's open road. Again, I'm about to close a case missing Grandma Helen Sarino last Sunday. Her grandson just reported she went out for bagels and never came back. This morning, we picked up an old lady matching her description. Exactly. I showed her pictures of Judd and she said, and I quote my grandson.

00:00:50
Speaker : What did I say about doing voices?

00:00:52
Speaker : I'm a storyteller, sir. It's my craft. Anyways, grandson's coming in. They reunite and I throw another case on the old solved it pile. Hey, Microsoft.

00:01:10
Speaker : You want to see me, Captain?

00:01:11
Speaker : Yes, the D.A. wanted me to personally thank you for your work on the Jane Street drug bust.

At the bottom of the interface, there is a "Saved a minute ago" notification, a "Realign timestamps" button, a "Show all shortcuts" button, a "Want audio to jump when clicking on text? Keep alt pressed" checkbox, and a "Replay Tour" link.

Implementation

- Determine what material is “ASR-processable”.
- Create the subtitles on the ASR platform.
- Conform the file.

Conclusions

- Use different ASR software.
- Decide whether to use the tool's editor or the usual subtitling software.
- Remember you are bound by NDAs.
- **Pricing models:** by the minute, per day, per user; pre-purchase batches of minutes, bundles, packages.
- The ideal post-editor.

The elusive “croissant”

- Out of the 22 ASRs we tested, only 4 transcribed this correctly.
- The rest understood “Hey, Microsoft.”
- Those 4 softwares scored above 70% in accuracy.



Fun with ASR!

- "WHERE DO I AFFIX THE PRINCESS TOWER, BOYLE?"
- where do I fit the prince's towel boy
- "I CAN'T HEAR YOU, SHIRT'S OVER MY EARS"
- I can't hear your shirts on my ears
- "CHARLES IS HOGGING MY DOLLHOUSE"
- Charles is hugging my daughter

Thank you!

- CI CE Code: ATA22LS43.
- Email: maracampbell@truesubtitles.com
- Twitter: [@truesubtitles](https://twitter.com/truesubtitles)

