

Live-Transcription Glasses Analogies Paper

Gabriela Knox

Communication & Journalism

COMM 205 503

Dr. Adam Bajan

March 5, 2025

Introduction

Live-transcription glasses, also known as live-captioning glasses, have the potential to be one of the most revolutionary tools in modern communications. Live-transcription glasses enable real-time transcription and translation of speech into text, using microphone, transcription, and display agents, to improve communication transactions for deaf or hard-of-hearing individuals, and breaking language barriers. Though the technology of live-transcription glasses may seem complex, this paper will cover two different analogies: running a relay race and ordering a pizza, to explain and simplify the technological process.

Analogy for Kids: Running a relay race

The technological process of how live-transcription glasses work is similar to the structure of a relay race, in which the baton acts as the data being transferred.

Point One : The microphone agent is like the first runner

The microphone agent is similar to the first leg of the relay race and acts as the first runner. As the microphone agent is responsible for picking up audio data sets through small, strategically placed microphones, it is the first runner's responsibility to pick up the baton and run the first lap (Yadava, et al.). After the microphones have picked up satisfactory audio data, it gets passed to the cloud infrastructure to begin the transcription process, similar to how the first runner finishes their lap to pass the baton to the second runner.

Point Two: The transcription agent is like the second runner

The second runner of the relay race acts as the transcription agent, in which the audio data is being translated from audio into speech. The middle portion of a relay race is generally the slowest, and the transcription agent has proven to be the most needed area of improvement due to slight transcription or translation delays (Yadava, et al.). Some relay racing strategies

include adding a stronger and faster runner in the middle, which can be related to the usage of Torch or Whisper for a quicker and more fluid transcription conversion. As the transcription is finished and passed through the cloud to the display agent, the middle runner passes the baton along to the last runner for the last leg of the race.

Point Three: The display agent is like the third runner crossing the finish line

The third and final runner of the relay race demonstrates how the transcribed text is displayed for the viewer, similar to the runner ending the race. Once the transcription is complete, the text data is sent to the display agent to be transmitted onto the OLED, which can navigate the text to ensure the correct text is being displayed within the context of the conversation or omit text as needed (Yadava, et al.). Once the text is displayed, the viewer will be able to read the fully transcribed text on the lens of the glasses. This is like how the third runner, and usually the quickest runner, is passed the baton to then run the final lap of the race. The runner can navigate and adjust speed or strategy to make up for lost time or placements throughout the race, and is responsible for ending the race strongly, but also for running across the finish line for the entire team.

Conclusion

In conclusion, similar to how a relay race involves parts of a team passing a baton through a race, the technological process of how live-transcription glasses work involves passing spoken words through different parts of a process to receive readable text.

Analogy for Pizza: Ordering a pizza

Ordering a pizza is similar to the process of how live-transcription glasses work, in which the process of a pizza order becoming a delivered pizza acts as spoken words becoming written text.

Point One : The microphone agent is like ordering a pizza

Ordering a pizza is similar to the process of the microphone agent in live-transcription glasses. When ordering a pizza over the phone, the person on the other side of the phone, or the pizza server, is listening carefully to the order. Similarly, microphones are strategically placed around the glasses, which pick up, distinguish, and isolate voices from background noises coming from different directions (Jaswanth, et al.). The reception of the order to the server demonstrates the reception of the audio from the microphones to the Raspberry pi zero W which processes the audio data and ensures seamless audio streaming (Jaswanth, et al.). The pizza server is able to listen to the order and distinguish between different choices to then note it down and process the order.

Point Two : The transcription agent is like assembling a pizza

The second aspect of live-transcription glasses is where the transcription occurs, and relates to the physical assembly of the pizza. The server relays the order to the chef in order to create the pizza, similar to how the audio data is relayed to the cloud-infrastructure to transcribe the speech into text. The glasses use different applications such as Whisper, an automated speech recognition system, or Torch, which offers a more seamless conversion, just like a chef might use different appliances to make the pizza such as an oven or pizza cutter (Jaswanth, et al.). Though baking a pizza is not instantaneous, once the pizza is fully cooked, it is sent back to the server to deliver it to the customer. This is similar to how the fully-transcribed text is sent back to the Raspberry pi zero w for final display after leaving the cloud-infrastructure system.

Point Three : The display agent is like the final delivery of a pizza

The display agent of live-transcription glasses acts similarly to the final delivery of the pizza to the customer. The OLED, or Organic Light Emitting Diode, display receives the

transcribed text from the Raspberry pi zero w, like how the customer receives the finished pizza from the pizza deliverer (Jaswanth, et al.) . The Raspberry pi zero w will ensure that the audio data is satisfactory and distinguishes whether or not certain information should be displayed, before sending the text to the display for the reader to see. Similarly, the pizza person will ensure that the pizza is satisfactory before boxing it, preparing it for delivery, and delivering it to the customer.

Conclusion

In conclusion, just how the three agents work together to listen, transcribe, and display speech to text, the process relates to the ordering, assembly, and delivery of a pizza. This analogy helps to break down the details of how live-transcription glasses work to provide a simpler understanding.

Bibliography

Fernando Loizides, Sara Basson, Dimitri Kanevsky, Olga Prilepova, Sagar Savla, and Susanna

Zaraysky. 2020. Breaking Boundaries with Live Transcribe: Expanding Use Cases

Beyond Standard Captioning Scenarios. In Proceedings of the 22nd International ACM

SIGACCESS Conference on Computers and Accessibility (ASSETS '20). Association

for Computing Machinery, New York, NY, USA, Article 12, 1–6.

<https://doi.org/10.1145/3373625.3417300>.

P. Jaswanth, C. H. Priya, K. Thota and M. Khanna, "Live Transcription and Closed Caption,"

2024. 15th International Conference on Computing Communication and Networking

Technologies (ICCCNT), Kamand, India, 2024,

<https://10.1109/ICCCNT61001.2024.10724562>.