



Video on Demand Toolkit: A Framework for Analysis of Speech and Chat Content in YouTube and Twitch Streams

Steven Coats

English, University of Oulu, Finland

steven.coats@oulu.fi

CMC-Corpora 11, University of Provence-Côte d'Azur

September 5th, 2024

Outline

1. Background
 - Video streams as an increasingly popular CMC modality
 - Corpus-based study of video streams
2. VoD Toolkit: Pipeline components
3. Use cases
 - Chat density
 - Automated analysis of video streams
4. Outlook and summary

Slides for the presentation are on my homepage at <https://cc.oulu.fi/~scoats>

Background

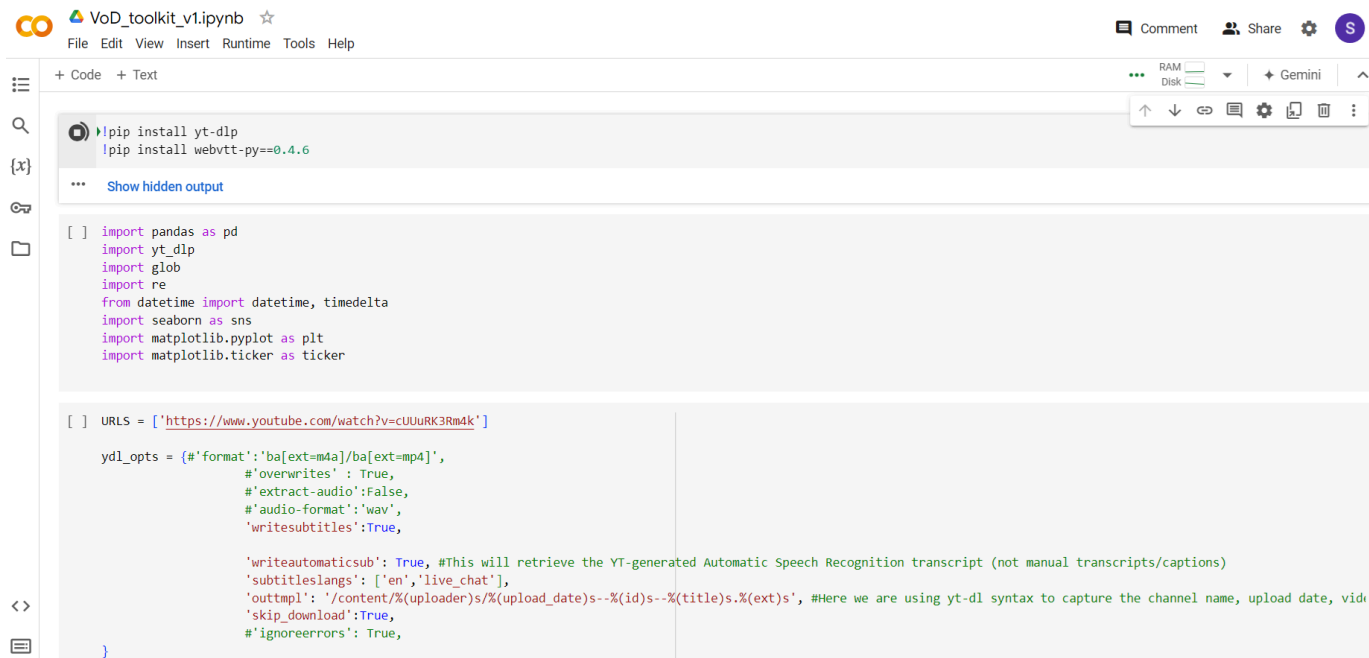
- In the past 15 years, video-based CMC modalities have become popular
- Many streaming sites have large numbers of users
 - Twitch (mostly gaming), YouTube Live (mostly music, live vlogging, tutorials, etc.), Instagram Live, Facebook Live, X Livestream, and others
 - Increasing importance as an economic activity (Zhou et al. 2019; Johnson & Woodcock 2019; Yu et al. 2018)
- Recorded streams contain multiple levels of communication at multiple levels (Sjöblom et al. 2019; Recktenwald 2017), for example
 - Speech and visual content (e.g. facial expressions or gestures) of the streamer
 - Text and graphical image content (emoji, emotes) of live chat
 - Text and graphical content of system messages (e.g. bots showing tips to streamer)
 - Secondary visual content (and text and speech) of video output (e.g., window showing gameplay)
- These offer new perspectives for the study of online interactional coherence (Herring 1999)
- Most corpus-based analyses have focused on live chat content (Olejniczak 2015; Kim et al. 2022)
- Few studies consider the content of the streamer as well as chat and comments

VoD Toolkit (https://t.ly/le6_e)

This contribution: A script pipeline to generate a structured, time-aligned transcript that combines the stream speech transcript with chat contributions and other types of content

- Python Jupyter environment
- Google Colab
- Generates textual output that can be analyzed with corpus methods
- Includes emoji and custom emotes
- Can also be used to capture video/audio for combined multimedia analysis

Pipeline components



```

VoD_toolkit_v1.ipynb
File Edit View Insert Runtime Tools Help
+ Code + Text
Comment Share
RAM Disk Gemini
... Show hidden output
[ ] import pandas as pd
import yt_dlp
import glob
import re
from datetime import datetime, timedelta
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker

[ ] URLs = ['https://www.youtube.com/watch?v=cUuRK3Rm4k']

yd_opts = {'format': 'ba[ext=m4a]/ba[ext=mp4]',
           #'overwrites': True,
           #'extract-audio': False,
           #'audio-format': 'wav',
           #'writesubtitles': True,

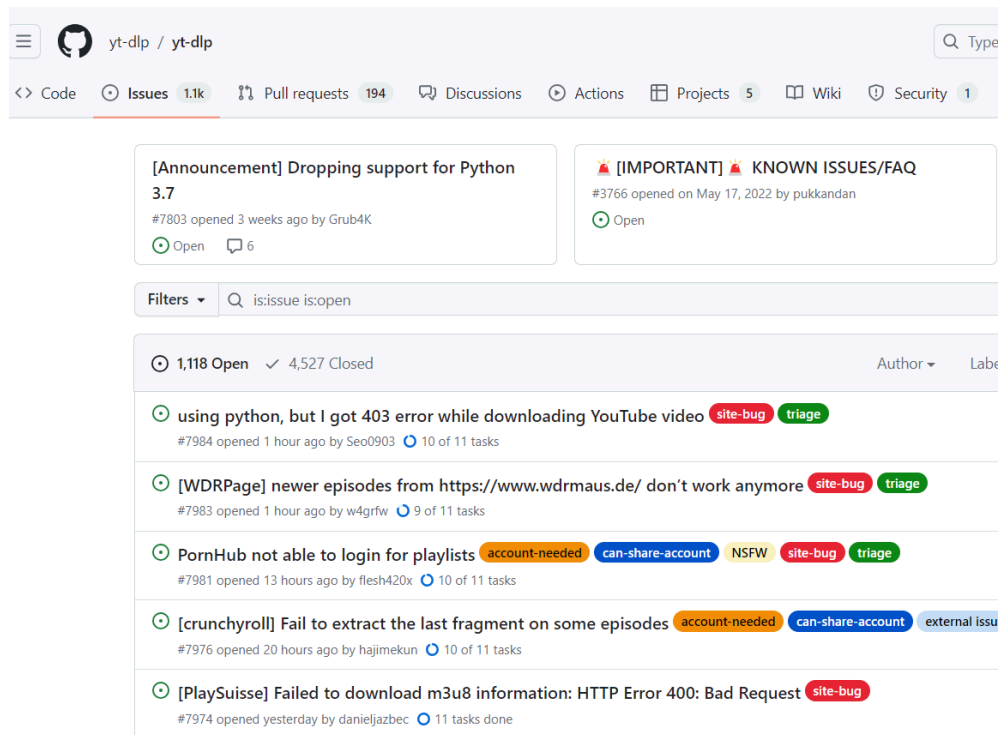
           #'writeautomaticsub': True, #This will retrieve the YT-generated Automatic Speech Recognition transcript (not manual transcripts/captions)
           #'subtitleslangs': ['en', 'live_chat'],
           #'outtmpl': '/content/(uploader)s/(upload_date)s--%(id)s--%(title)s.%(ext)s', #Here we are using yt-dl syntax to capture the channel name, upload date, vid
           #'skip_download': True,
           #'ignoreerrors': True,
           }

```

- yt-dlp
- TwitchDownloaderCLI
- Whisper/WhisperX

The toolkit's output is an HTML file

Component: yt-dlp



yt-dlp / yt-dlp

<> Code Issues 1.1k Pull requests 194 Discussions Actions Projects 5 Wiki Security 1

[Announcement] Dropping support for Python 3.7
#7803 opened 3 weeks ago by Grub4K
Open 6

🔥 [IMPORTANT] 🔥 KNOWN ISSUES/FAQ
#3766 opened on May 17, 2022 by pukkanan
Open

Filters is:issue is:open

1,118 Open 4,527 Closed Author Labels

- using python, but I got 403 error while downloading YouTube video **site-bug** **triage**
#7984 opened 1 hour ago by Seo0903 10 of 11 tasks
- [WDRPage] newer episodes from https://www.wdrmaus.de/ don't work anymore **site-bug** **triage**
#7983 opened 1 hour ago by w4grfw 9 of 11 tasks
- PornHub not able to login for playlists **account-needed** **can-share-account** **NSFW** **site-bug** **triage**
#7981 opened 13 hours ago by flesh420x 10 of 11 tasks
- [crunchyroll] Fail to extract the last fragment on some episodes **account-needed** **can-share-account** **external issue**
#7976 opened 20 hours ago by hajimekun 10 of 11 tasks
- [PlaySuisse] Failed to download m3u8 information: HTTP Error 400: Bad Request **site-bug**
#7974 opened yesterday by danieljazbec 11 tasks done

- Fork of YouTube-DL
- Can be used to access any content streamed with DASH or HLS protocols
- Can be used to get video

Component: TwitchDownloaderCLI



Twitch Downloader

Twitch VOD/Clip/Chat Downloader and Chat Renderer

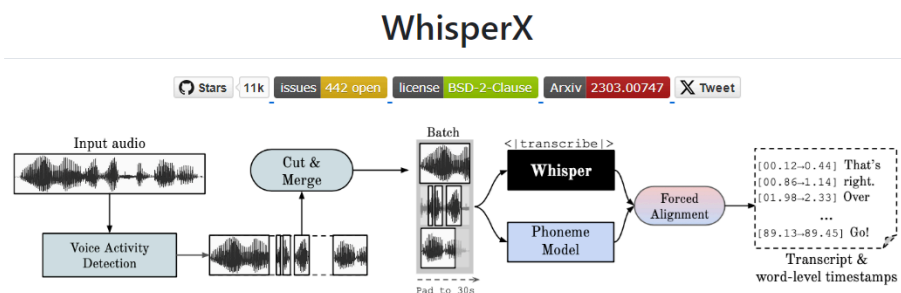
[Report Bug](#)

This document is also available in:

- [Spanish / Español](#)
- [Italian / Italiano](#)
- [Portuguese \(Brazil\) / Português \(Brasil\)](#)
- [Turkish / Türkçe](#)
- [Japanese / 日本語](#)
- [Simplified Chinese / 简体中文](#)
- [Russian / Русский](#)

- Command-line interface for retrieving Twitch videos and chats

Component: WhisperX



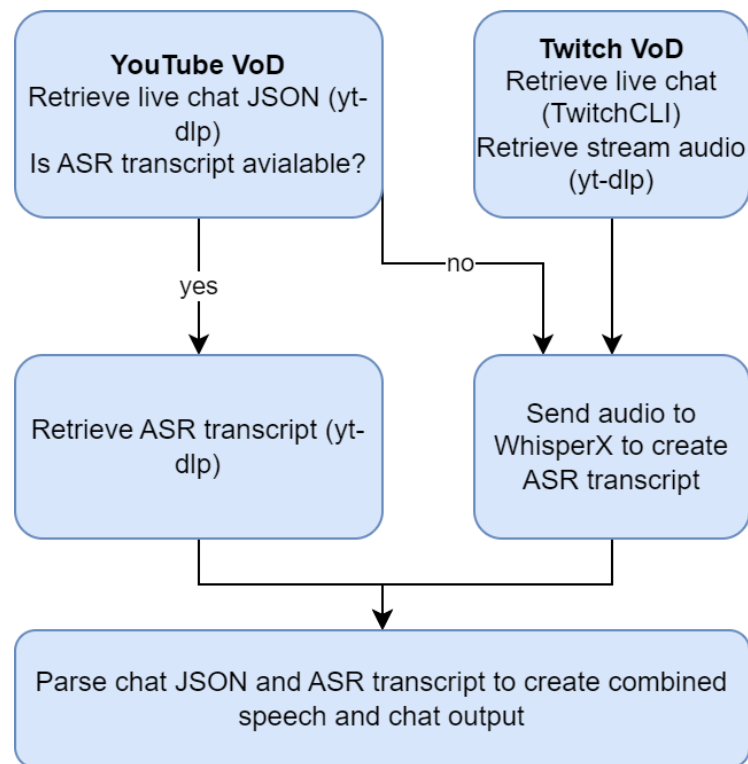
This repository provides fast automatic speech recognition (70x realtime with large-v2) with word-level timestamps and speaker diarization.

- ⚡ Batched inference for 70x realtime transcription using whisper large-v2
- 🔥 [faster-whisper](#) backend, requires <8GB gpu memory for large-v2 with beam_size=5
- 🎯 Accurate word-level timestamps using wav2vec2 alignment
- 👤 Multispeaker ASR using speaker diarization from [pyannote-audio](#) (speaker ID labels)
- 🔊 VAD preprocessing, reduces hallucination & batching with no WER degradation

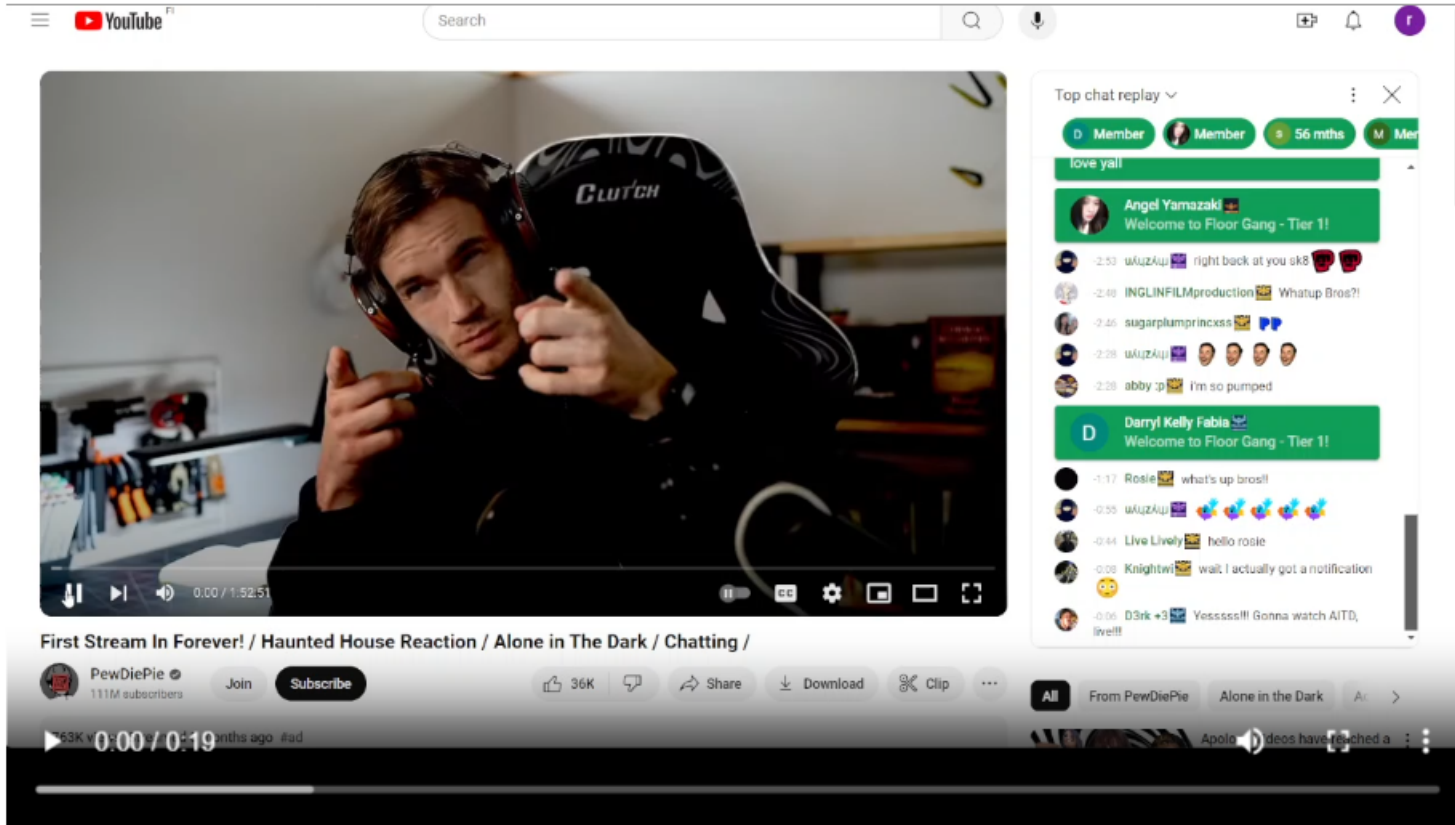
Library based on OpenAI's Whisper providing Automatic speech recognition

- Word-level timestamps
- Speaker diarization
- Faster than Whisper, especially with GPU

Schematic representation of pipeline functions



Example: YouTube stream



The screenshot displays a YouTube live stream interface. The main video player shows a streamer wearing a headset and holding a controller, with a 'GLITCH' logo on the chair. The video title is 'First Stream In Forever! / Haunted House Reaction / Alone in The Dark / Chatting /'. The streamer's name is 'PewDiePie' with 111M subscribers. The video has 36K likes. The chat window on the right shows a 'Top chat replay' with messages from users like 'Angel Yamazaki' and 'Darryl Kelly Fabia'.

YouTube

Search

Top chat replay

Member Member 56 mths M Mer

love yall

Angel Yamazaki Welcome to Floor Gang - Tier 1!

-2:53 u4uzAuj right back at you sk8

-2:48 INGLINFILMproduction Whatup Bros?!

-2:46 sugarplamprincekx

-2:28 u4uzAuj

-2:28 abby 3p I'm so pumped

Darryl Kelly Fabia Welcome to Floor Gang - Tier 1!

-1:17 Rosie what's up brosi!

-0:58 u4uzAuj

-0:44 Live Lively hello rosie

0:00 Knightw wait I actually got a notification

0:00 D3rk +3 Yessss!!! Gonna watch AITD, live!!!










First Stream In Forever! / Haunted House Reaction / Alone in The Dark / Chatting /

PewDiePie 111M subscribers Join Subscribe

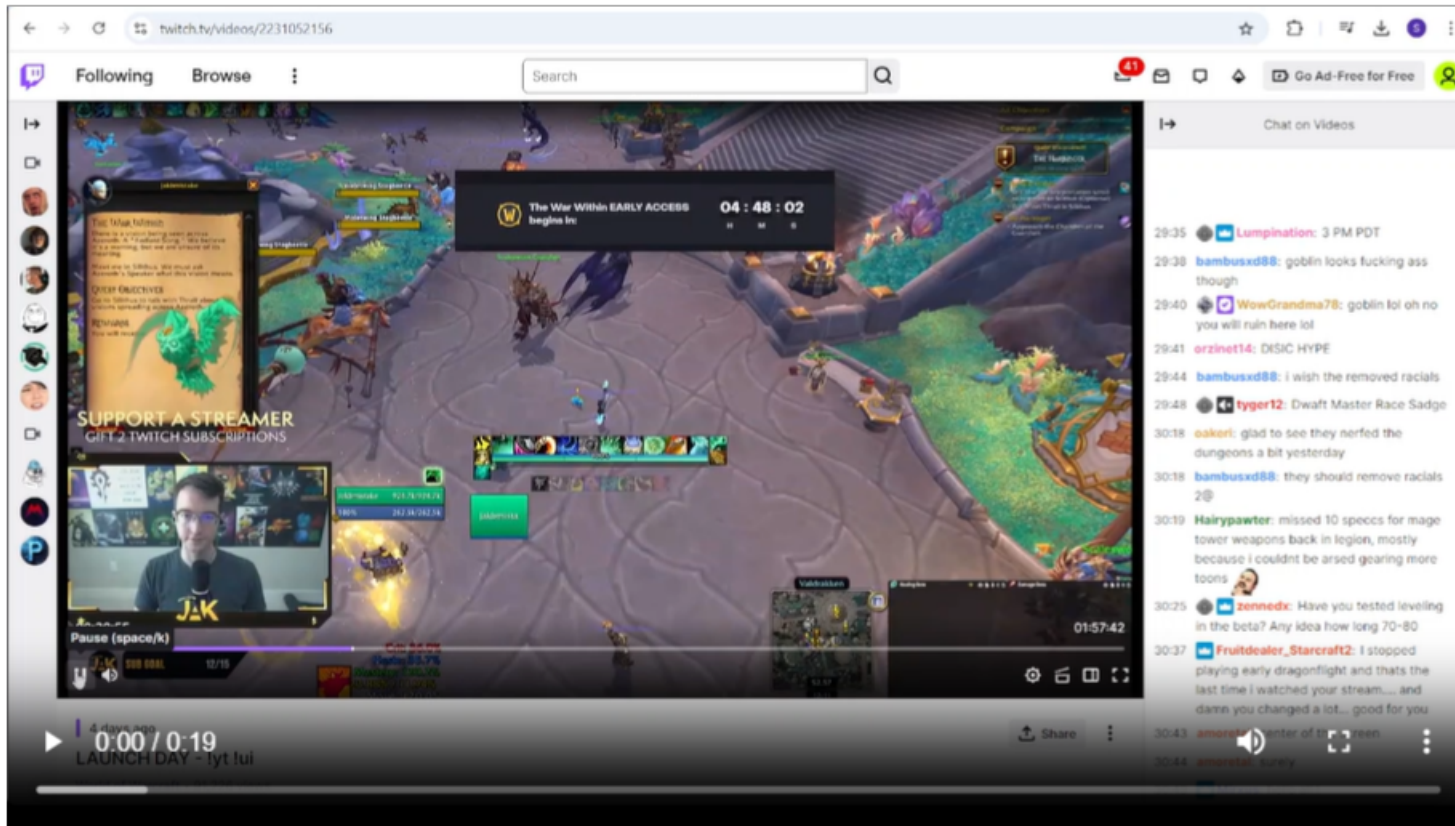
36K Like Comment Share Download Clip

63K views 0:00 / 0:19








Example: YouTube output

time	text	author	message
-1429.0		j 	you mean play?
-1359.0		Blank Budder 	well it has #ad in the description. I don't mind though, the haunted house video was great
-1350.0		ScottK 	
-1348.0		ScottK 	
-1346.0		ScottK 	
-1323.0		ScottK 	

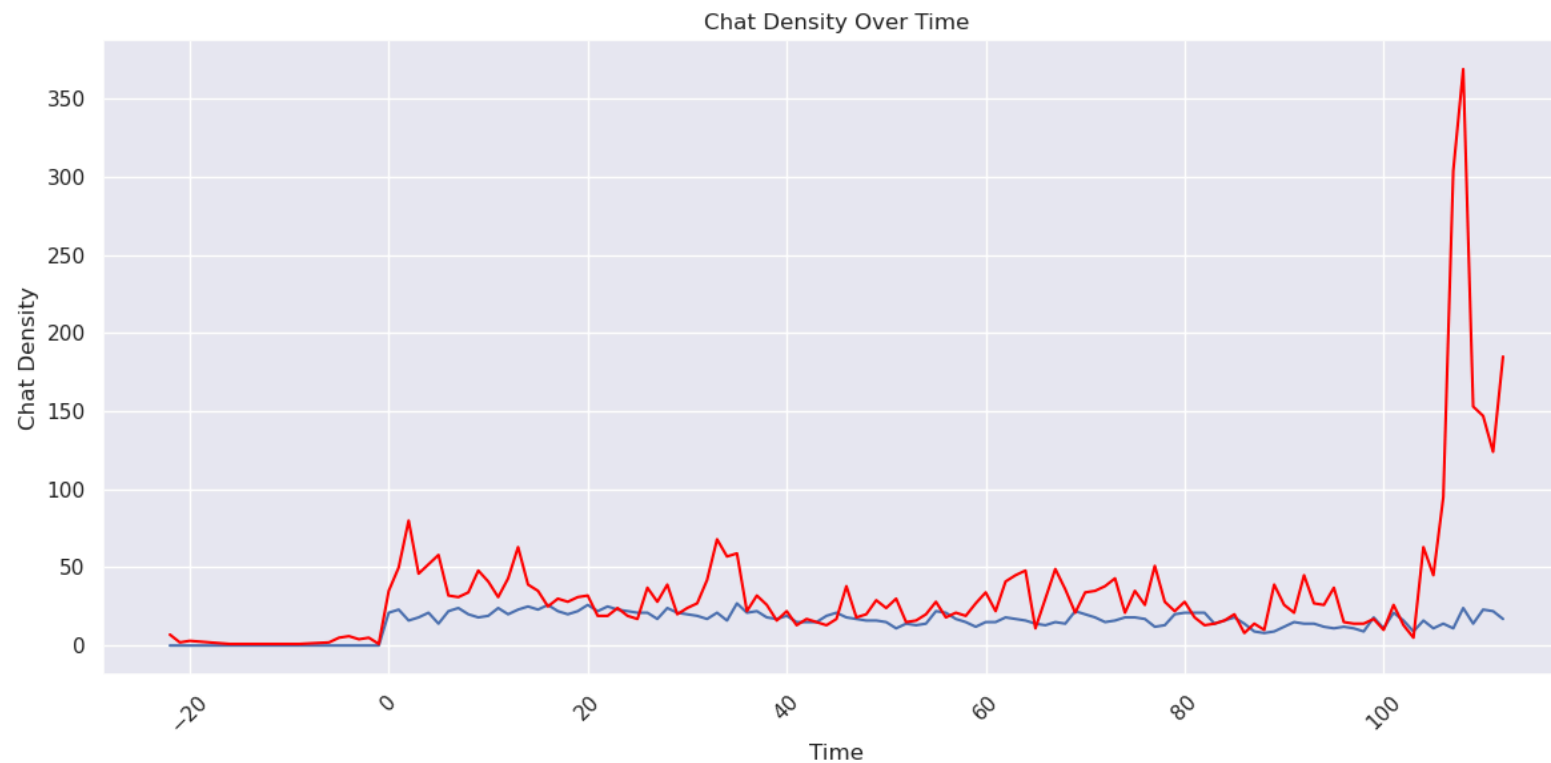
Example: Twitch stream



Example: Twitch output

time	text	author	message
11.000		StreamElements 	AutomaticJak is now live! Streaming World of Warcraft: LAUNCH DAY - !yt !ui 
28.000		zanis_	We here!
39.000		Goldenhusk	LICKA 
42.000		lissargh 	wooh hype launch day launch day launch day
77.000		CarlCYuro 	Woohoo got my snacks, foods, dirnks, dinner meal prepped for the next 3 days lets go
108.000		teddyx87 	lol
118.000		teddyx87 	what you maining Carl?

Use cases: Chat density

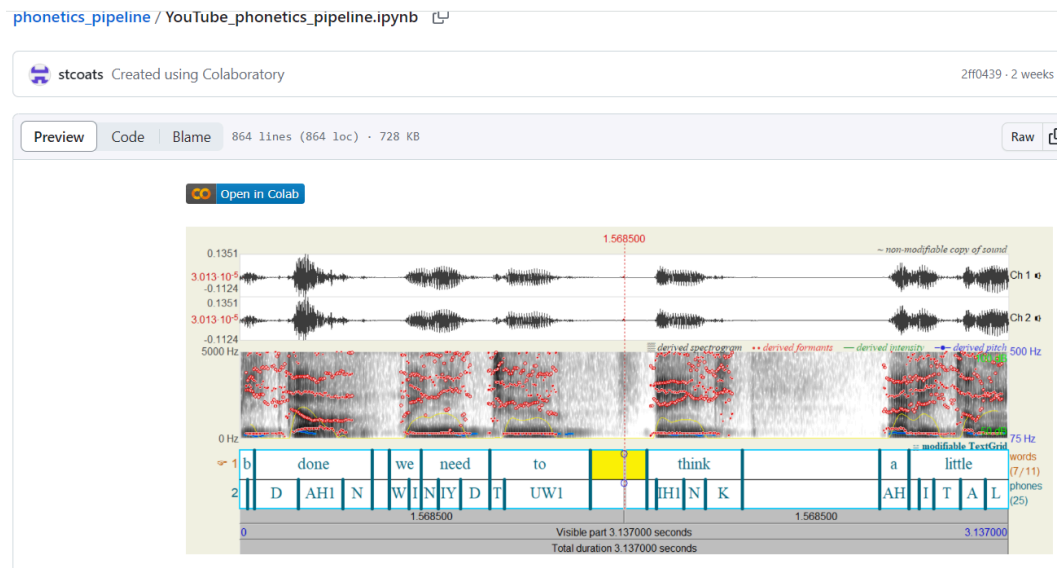


- Chat density can be compared with and correlated with streamer utterances

Potential use case: Automated analysis of video streams

- Retrieve video with yt-dlp
- Add cells to VoD Toolkit to import (e.g.) **X-CLIP** (Ni et al. 2022), **LLaVA-NeXT-Video-7B-h9** (Zhang et al. 2022), or other libraries
 - Automatically generate text describing what is going on in different parts of the video
 - Who is chatting about what parts of the video?
 - Is chat about (for example) video content, other chat, or speech content?

Use cases: Acoustic analysis



- Acoustic features of particular streamers or streams with different topics/from different locations etc. can be analyzed (cf. Coats 2023; Méli et al. 2023)

https://colab.research.google.com/github/stcoats/phonetics_pipeline/blob/main/phonetics_pipeline_v3.ipynb

Summary and outlook

- Ready-to-use scripting pipeline for generation of combined speech transcript/live chat files
- Can be used for various text-based corpus-analytic research questions
- Can be used, with additional modules, for multimodal analysis of video and audio content

Thank you for your attention!

References

- Coats, S. (2023). [A pipeline for the large-scale acoustic analysis of streamed content](#). In Louis Cotgrove, Laura Herzberg, Harald Lungen, and Ines Pisetta (eds.), *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2023)*, 51–54. Mannheim: Leibniz-Institut für Deutsche Sprache.
- Herring, S. (1999). [Interactional coherence in CMC](#). *Journal of Computer-Mediated-Communication*, 4(4).
- Johnson, M. R., & Woodcock, J. (2019). [The impacts of live streaming and Twitch.tv on the video game industry](#). *Media, Culture & Society*, 41(5), 670–688.
- Kim, J., Wohn, D. Y., & Cha, M. (2022). [Understanding and identifying the use of emotes in toxic chat on Twitch](#). *Online Social Networks and Media* 27.
- Méli, Adrien, Steven Coats and Nicolas Ballier. (2023). [Methods for phonetic scraping of Youtube videos](#). In *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)*, 244–249.
- Ni, B., Peng, H., Chen, M., Zhang, S., Meng, G., Fu, J., Xiang, S., & Ling, H. (2022). [Expanding language-image pretrained models for general video recognition](#). *arXiv*, cs.CV, 2208.02816.
- Olejniczak, J. (2015). A linguistic study of language variety used on twitch.tv: Descriptive and corpus-based approaches. In *Proceedings of RCIC'15: Redefining Community in Intercultural Context, Brasov, 21–23 May 2015* (pp. 329–334).
- Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics* 115, 68–81.
- Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2019). The ingredients of Twitch streaming: Affordances of game streams. *Computers in Human Behavior*, 92, 20–28.
- Yu, E., Jung, C., Kim, H., & Jung, J. (2018). Impact of viewer engagement on gift-giving in live video streaming. *Telematics and Informatics*, 35(5), 1450–1460.
- Zhang, Y. Li, B., Liu, H., Lee, Y. G., Gui, L., Fu, D., Feng, J., Liu, Z., & Li, C. (2024). [LLaVA-NeXT: A Strong Zero-shot Video Understanding Model](#).
- Zhou, J., Zhou, J., Ding, Y., & Wang, H. (2019). The magic of danmaku: A social interaction perspective of gift sending on live streaming platforms. *Electronic Commerce Research and Applications* 34, 100815.