Disagreement Detection

Harvard John A. Paulson School of Engineering and Applied Sciences

Spotify®

Milestone 1

Connor Capitolo, Max Li, Morris Reeves, Jiahui Tang

Overview

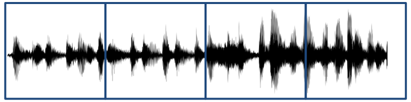
- Problem statement and scope of work
- Infrastructure and annotation progress
- Literature review and current ideas
- EDA:
  - Metadata
  - Audio
  - Text
- Next Steps
- Q&A

What are we trying to achieve?

Generate insights about disagreement based on Spotify podcast data

Inputs

Audio, text, and metadata



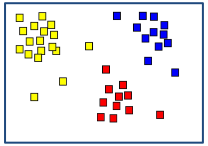
No, that's absurd..	Ok, moving on...	This is a good podcast	No, it's been quite bad...

Outputs

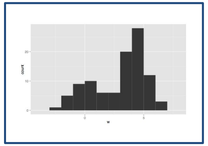
Disagreement/Non-disagreement

D	N	N	D

Insights



Clusters:
What genres have high predicted disagreement?



Time:
When are periods of predicted disagreement?

Harvard John A. Paulson School of Engineering and Applied Sciences

Why is disagreement detection in podcasts important & interesting?

- **Higher Retention**
- Improved Recommender
- Expand platform ecosystem

- **Better User Experience**
  - Tailored recommendations

- **Drive audience traffic + loyalty**
- Improve style or moderation

Infrastructure

Communication

slack

WhatsApp

Data/Code Storage

GitHub

box

aws

Packages

librosa

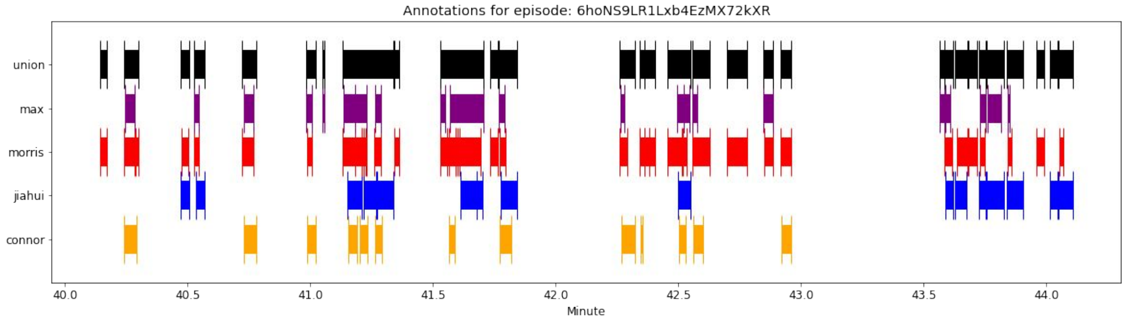
pandas

prodigy

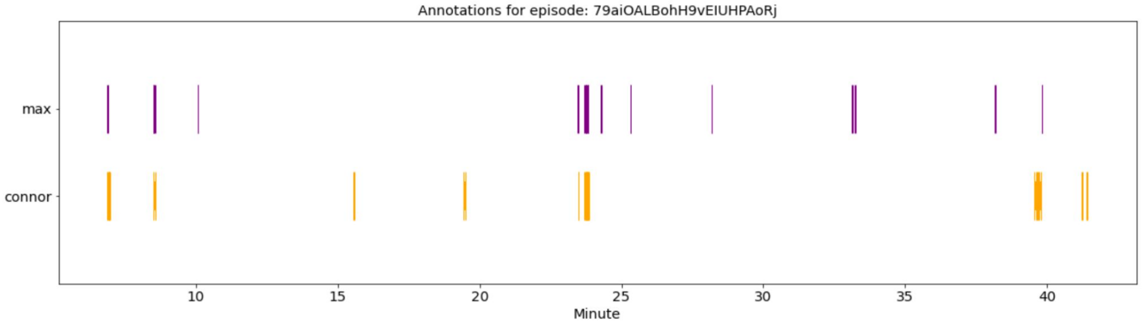
Annotation progress

- Working definition of disagreement: a speaker is **directly applying a contradiction or rejection** of another person's idea where it is **immediately perceptible** by the listener

Dog clip



Full episode



Literature

2. Primarily utterance-based granularity

	Xu et al. (2019)	Gokcen and de Marneffe (2015)	Wang and Cardie (2014)	Wang et al. (2011)	Hillard et al. (2003)
Disagreement definition or focus	Independent, differing stance-bearing *utterances*	*Quote-response* pairs; polarity or modality mismatch	*Utterance* or turn; “user’s attitude”	*Utterance*; speaker rejects proposition by a first speaker	*Spurt* (no pauses > 0.5 second); audio and text-based cues
Model class	“RCN” (Reason Comparing Network): NN	Logistic regression (max entropy model, Stanford CoreNLP)	Isotonic CRF	Linear-chain CRF	Decision tree classifier
Features	Utterance pair (P, Q) Topic T	N-grams, speech acts, typed dependencies, etc.	Dependency relations, TFIDF, n-grams, etc.	Lexical features, prosodic features (pause, duration, speech rate, pitch)	Word-based features, prosodic features (pause, fundamental frequency (F0), etc.)
Performance	~**0.58-0.73** macro F1	**0.76** F1	**0.51** F1	**0.56** F1	0.53 - 0.58 *Recall*
Data	**Text**: tweets (SemEval-2016, stance labels)	**Text:** forum (Internet Argument corpus)	**Text:** forum (Wikipedia Discussions corpus)	**Transcript + audio** (Broadcast Conversations)	**Transcript + audio** (ICSI Meeting transcript corpus)

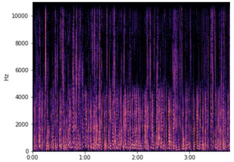
3. Performance is ~0.5 to 0.75 F1

1. Primarily text-based or highly structured data

Project ideas

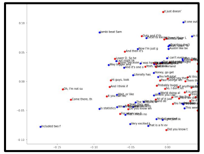
Current presentation

Audio featurization



- Prosodic features (e.g. pause, pitch, amplitude)
- Diarization

Transcript featurization



- Token-based embeddings
- Contextualized embeddings

Supervised modeling

Baseline

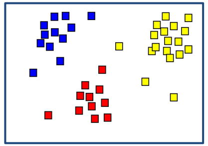
- Logistic regression

Sequential models

- RNN

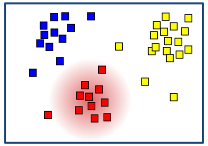
Unsupervised modeling

Audio + transcript clustering



- Join on timestamp
- Diarization-based segments

Similarity-based classification



- Based on distance to annotated disagreement examples

Harvard John A. Paulson School of Engineering and Applied Sciences

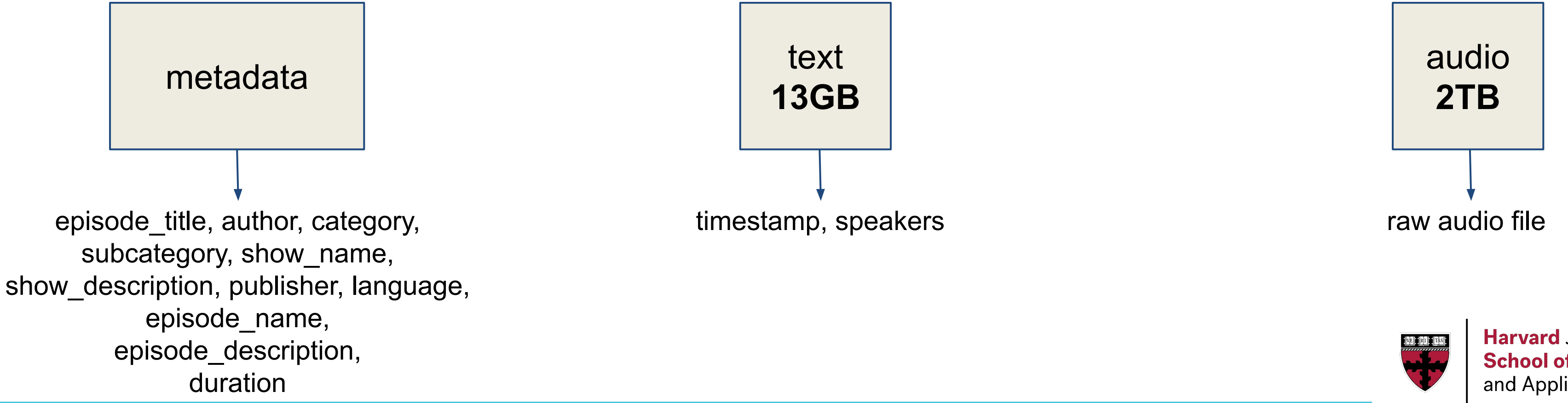
Metadata - Overview

105,360 episodes, 18,360 shows

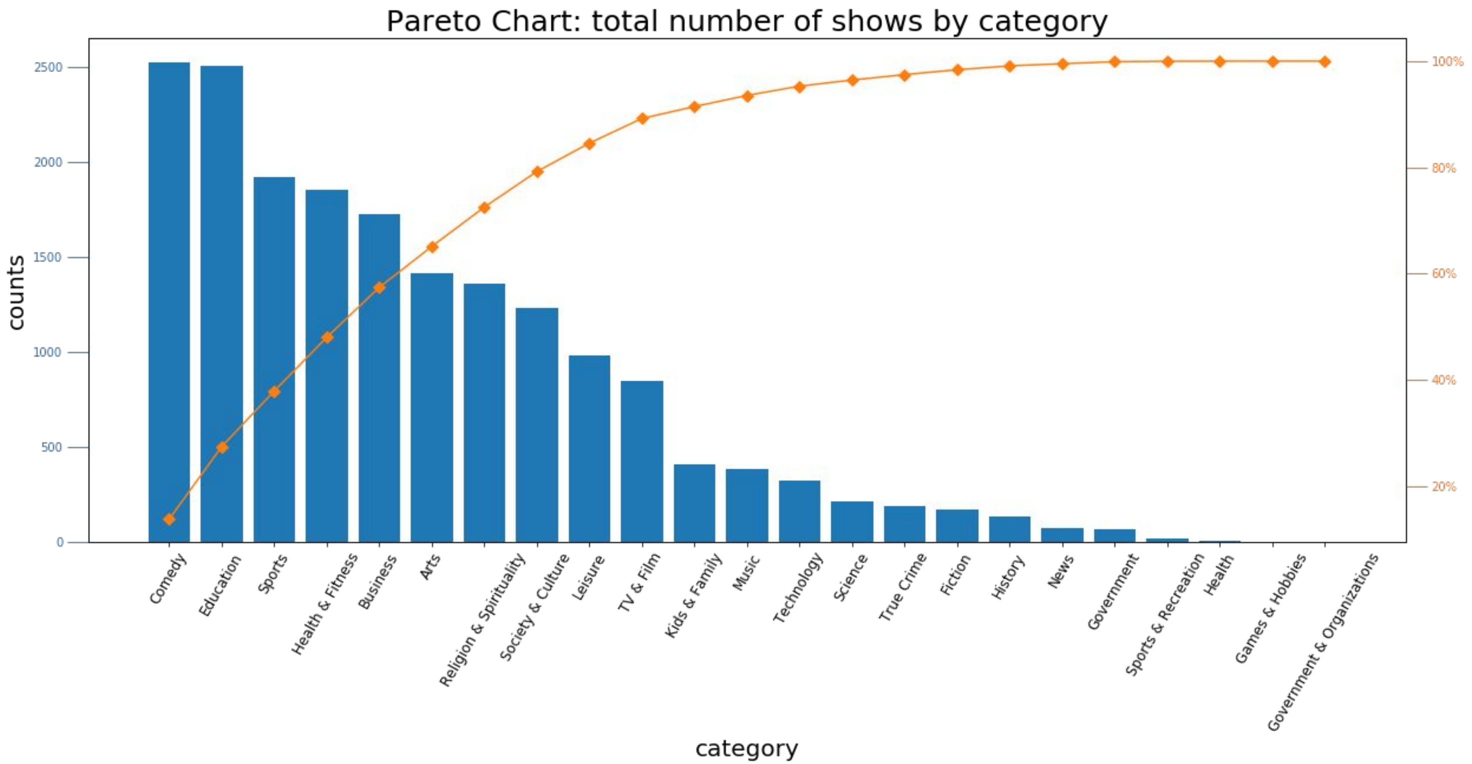
- ~ 50k hours of audio
- > 600M words
- Jan 1, 2019 to March 1, 2020
- English language specific
- Professional + amateur creators

Dataset statistics

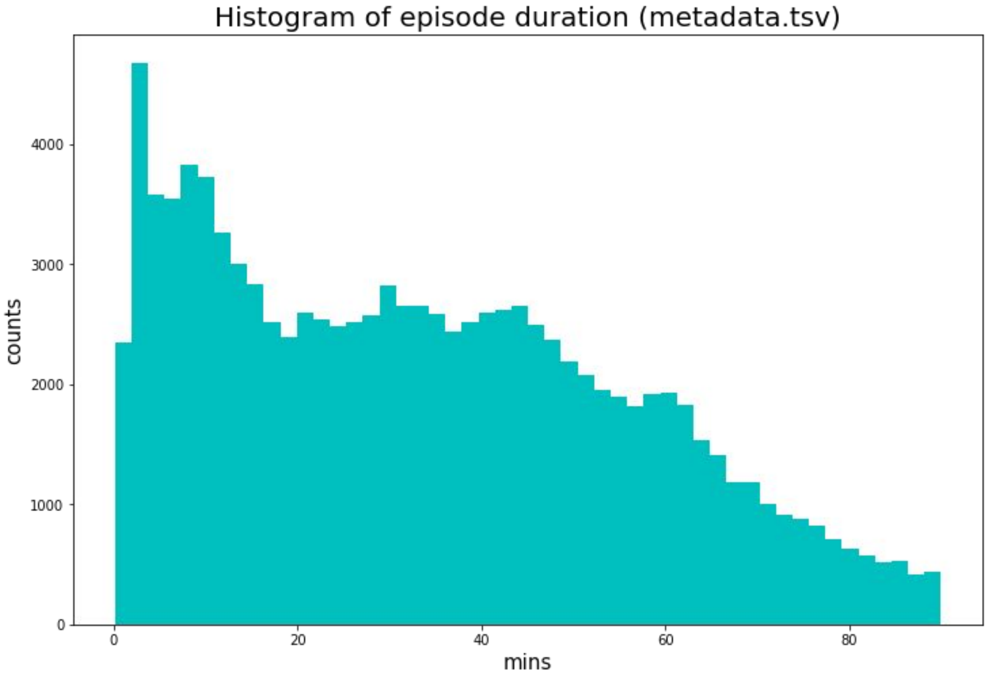
	min	average	max
minutes	<1	31.6	305.0
words	11	5,728	43,504



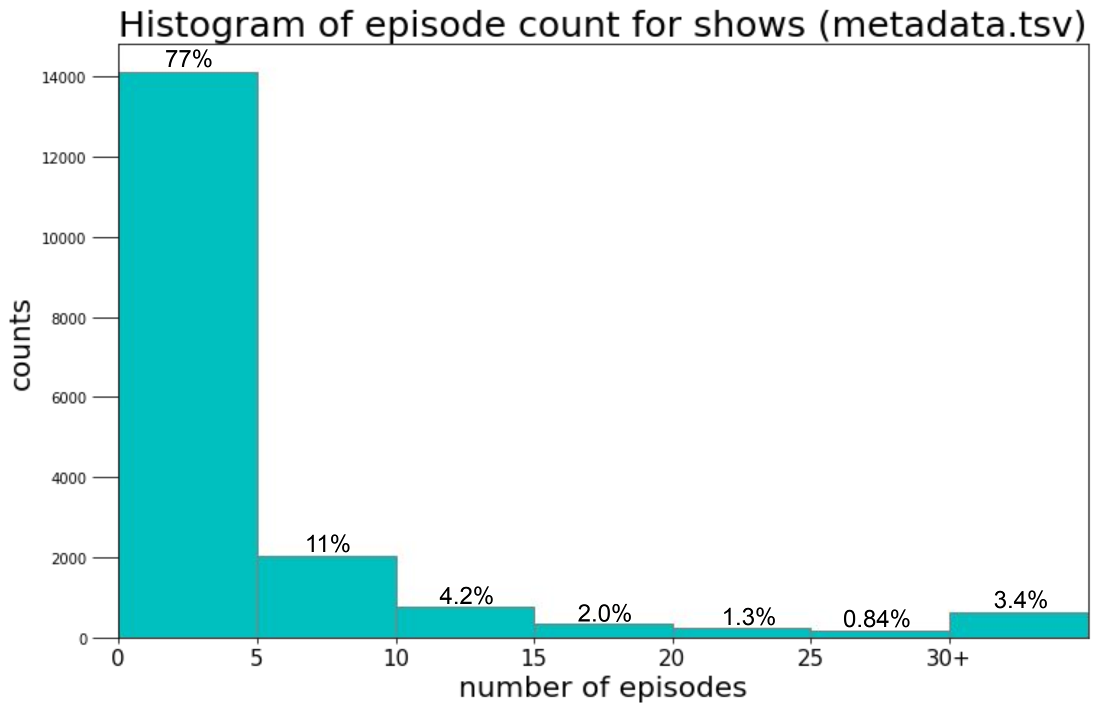
Metadata - genre



Metadata - episode duration

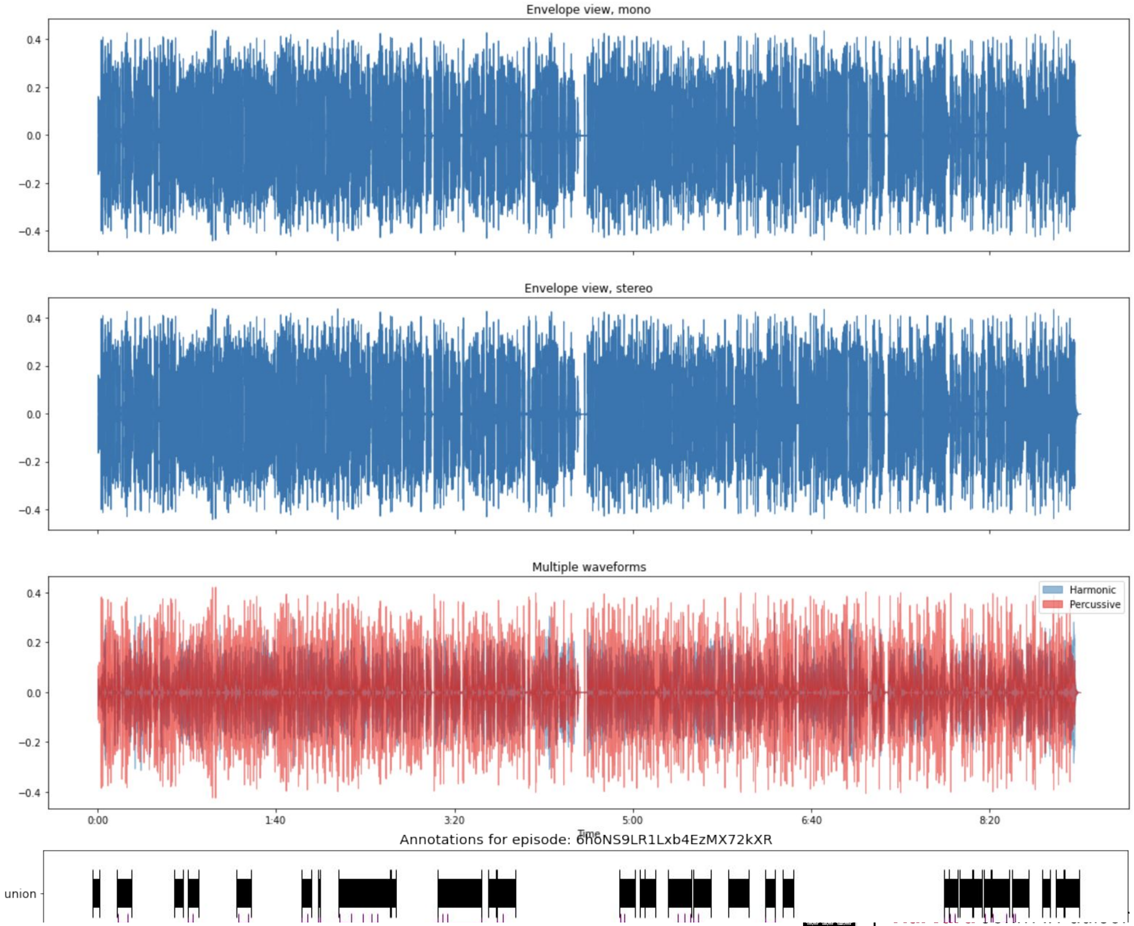


Metadata - number of episodes per show



Audio - Librosa

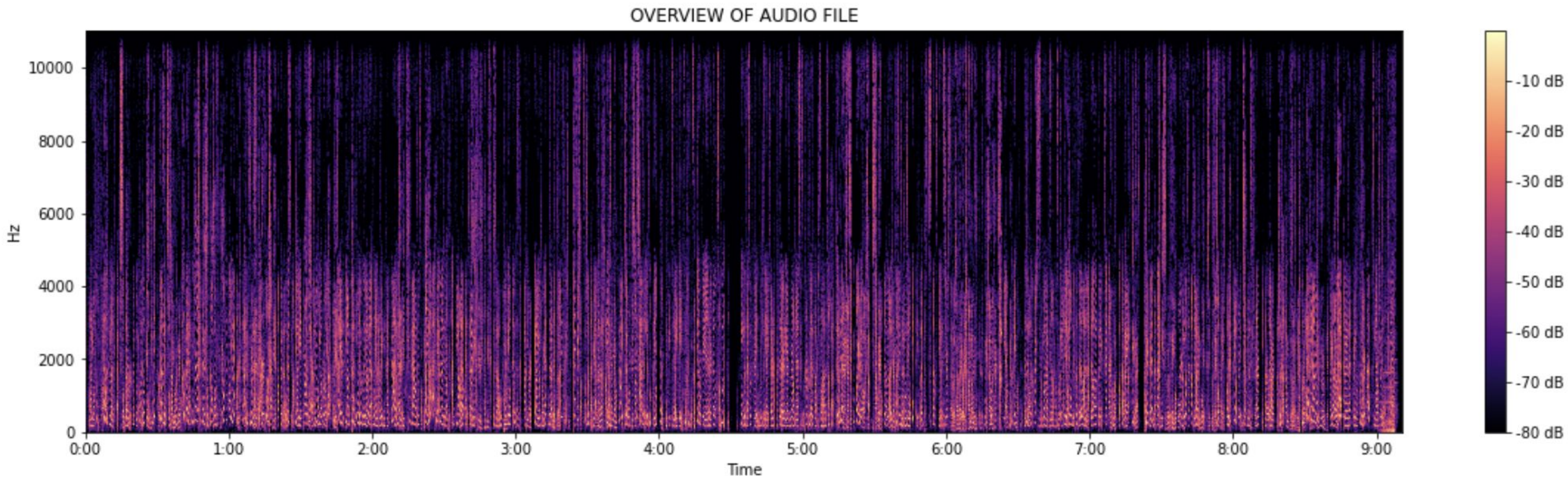
- Librosa is a Python package for **music and audio** analysis.
- For **single annotated episode**
  - ~10min Annotated Episode w.r.t Dog
- WaveForm (amplitude envelope)
  - Monophonic
  - Stereo
  - Harmonic + Percussive Components



School of Engineering
and Applied Sciences

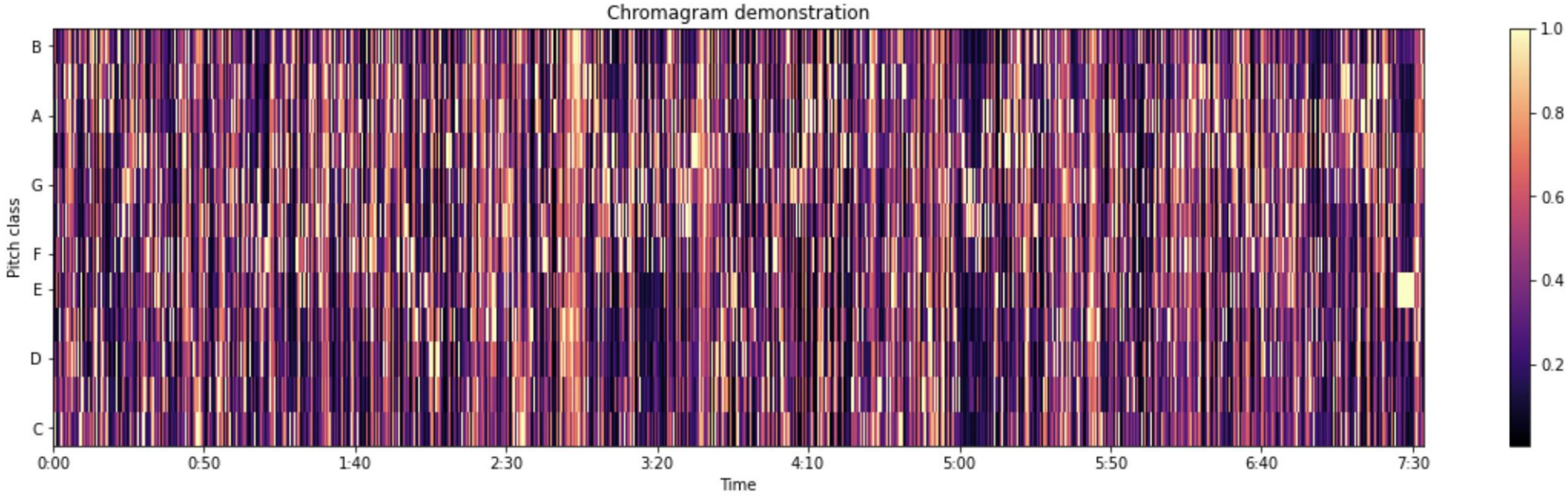
Spectrogram (Frequency)

It uses Short-time Fourier transform (STFT) to calculate signals in the **time-frequency(Hertz)** by computing discrete Fourier transforms over short overlapping windows



Chroma Plot (Pitch Class)

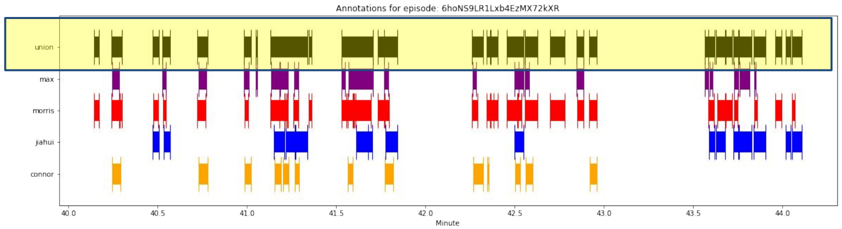
Chroma Plot gives information about Pitch Class Profile and its intensity, which is a powerful tool for analyzing music whose pitches can be meaningfully categorized

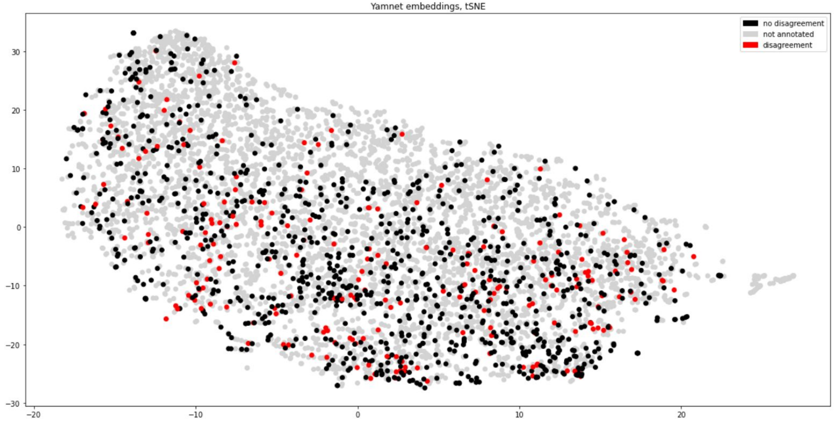


Audio - YAMNet

For single annotated episode...

- **Unioned** the disagreement annotations
- **YAMNet** - deep net that predicts 521 audio event classes (trained on AudioSet-YouTube corpus)
  - Mobilenet_v1 architecture
- Retrieved **Yamnet embedding** for episode:
  - Embeddings: 0.48 sec chunks, embedding dimension 1024
- Plotted tSNE of Yamnet embeddings vs. disagreement annotations





Harvard John A. Paulson School of Engineering and Applied Sciences

Text EDA Overview

- Creation of `Transcript` and `WordEmbeddingVectorizer` classes
  - Streamlined parsing of .json transcripts by using class instances
  - Allows flexibility in future swapping of text vectorization or episode discretization

```
gnews_vectorizer = WordEmbeddingVectorizer(embedding_dict = gnews_dict,
                                           embedding_dim = 300,
                                           lemmatize = False, remove_stopwords = True, n_token_filter = 10)
```

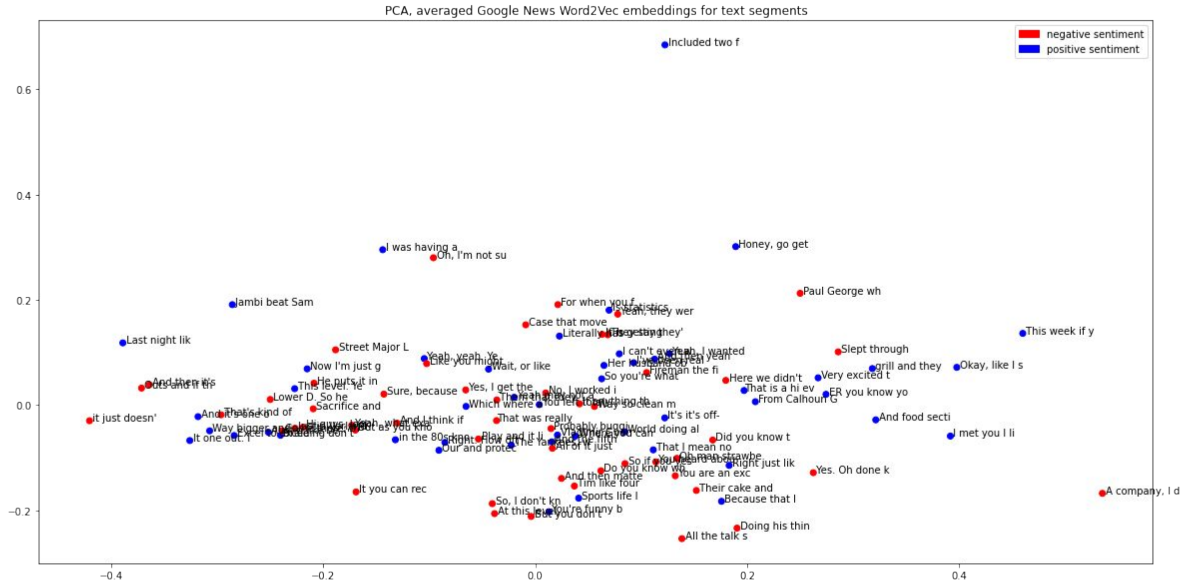
```
TRANSCRIPT_DIRECTORY = '../podcasts-no-audio-13GB/spotify-podcasts-2020/podcasts-transcripts/'
EXAMPLE_JSON_FILEPATH = TRANSCRIPT_DIRECTORY + '4/9/show_49NxrBHUtto19pgLNAJkHY/6hoNS9LR1Lxb4EzMX72kXR.json'
```

```
example_transcript = Transcript(EXAMPLE_JSON_FILEPATH, gnews_vectorizer, segment_definition = 'default')
example_transcript.set_segment_vectors(gnews_vectorizer)
example_transcript.set_sentiments(classifier)
```

- EDA:
  - Pretrained sentiment models on segments
  - Averaged (fixed) word embeddings

Text EDA: Sentiment

- Averaged word vectors vs. sentiment
- Exploration 1: **Sentiment and word embeddings** (for single podcast episode)



Harvard John A. Paulson School of Engineering and Applied Sciences

Text EDA: Baseline Model

- Baseline cosine similarity model for disagreement detection:
- Implemented as method of `Transcript` class

Cosine similarity threshold

```
example_transcript.get_disagreement_cosine("No, I completely disagree with you.", 0.7)
```

```
(array([" Doing his thing. There were Michael Jordan was just coming up there weren't there weren't many things to do that. You actually went and played outside. It was yeah, it was a different time. I feel like if people like it now of kids like it now is because they play Big play the score I think so not there's not a lot of or you know their families obsessed with it, you know, I don't know. No, I think that's I think that's true. But I think the biggest problem that baseball has is that they don't have a villain because I think every I mean you have",
       " If anything that is like not slightly Gary, you know that I'm right they keep a dog in this little tiny ass apartment. They never take it for walks because they're alone and can't find a boyfriend. It's not the dog's fault. You're a fucking loser. That is not first of all, are you are you saying? That's what I am. No, you're just laughing because he knows it's true. No, it's just the most ridiculous tape.",
       " You are an exception. Most people get animals for selfish reasons. It's not an exception though. I'm just an average normal person. Like I'm not an exceptionally not. Oh my gosh. I'm talking you up, but you're talking other women down and I totally disagree with them not talking them Dad. I'm telling the truth. No, I don't think that just like first of all you do get a dog because you want company. That's the only that's why people like what everybody gets dogs.",
       " Because that loud say what you're feeling is there's two of us. Yes, but when there was just one of us, which there was for many years of his life right out of his mind. Oh get over yourself. I love that. You think that you just add so much. I like that. You just can never admit that I'm right because that's not right you there's parts of it that truth is you're the one that own the dog so you don't want to see the other side. I said, I don't want to see what I've seen definitely got a dog because I wanted",
       " Did you know that happened? I heard that yeah, that's your to 2019 Mets season in a nutshell can wean you put people in the memoriam who are totally alive since I know I've how do you screw that up? That is eight easy Google search on baseball-reference very very simple. I mean, obviously we know if we know anything about the Mets is that they don't do their research. No, it's my clearly when it comes to scouting or"],
      dtype='<U642'),
 array([[ 539.8,  569.6],
        [2443.4, 2471.1],
        [2532.2, 2560.9],
        [2604.8, 2634. ],
        [2716.2, 2743.9]]),
 array([0.70385551, 0.70334772, 0.72889266, 0.70259684, 0.70927119]))
```

[Start, end] timestamps

Cosine similarity scores

Next steps

- Text-audio data join
  - Clustering on combination of text and audio features
- Smarter discretization of transcript and audio
  - e.g. pyannotate-audio diarization
- Scaling up to audio corpus and text corpus
  - Active learning
- Further disagreement annotations and refinement
- Baseline model and evaluation

Q&A

Thank You