Utrecht University

THE 7TH WORKSHOP ON ANALYSIS, UNDERSTANDING AND PROMOTION OF HERITAGE CONTENTS

Listening to Oral History:
Emotion Annotation and Recognition in the ACT UP Oral History Project

Francisca Pessanha, Ian Padovani, Justus van Klaveren, Heysem Kaya, Almila Akdag and Judith Masthoff

Utrecht University

The "deep dark secret" of oral history, is that nobody spends much time listening to or watching recorded and collected interview documents.

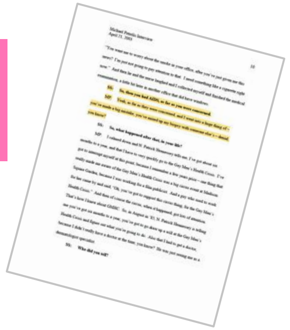
Michael Frisch

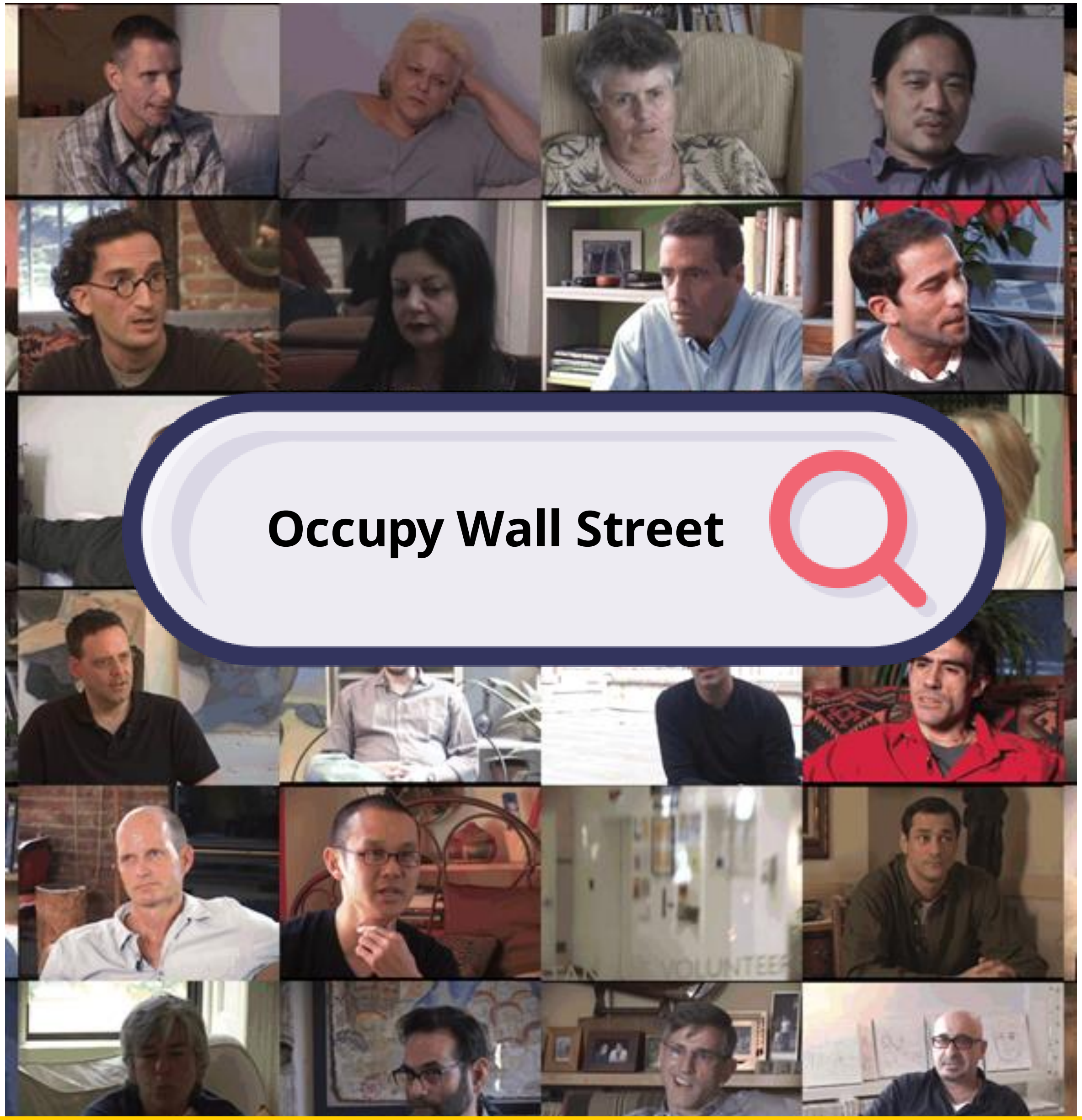
Utrecht University

Michael Frisch. 2008. Three dimensions and more: Oral history beyond the paradoxes of method. Handbook of emergent methods (2008), 221–238.

Computational techniques have brought new insights to the field of oral history, with methods such as:

Automatic Speech Recognition
and
Natural Language Processing.





Utrecht University

Francisca Pessanha and Almila Akdag Salah. 2021. A Computational Look at Oral History Archives. J. Comput. Cult. Herit. 15, 1.

The interview process is “a performance in search of a text”.

Reading only the interview transcriptions results in missing the performance itself.

Utrecht University

Siobhán McHugh. 2015. The affective power of sound: Oral history on radio. In The oral history reader. Routledge, 490–507.

Michael Petrelis Interview
April 21, 2003

"You want me to worry about the smoke in your office, after you've news? I'm just not going to pay attention to that. I need something now." And then he and the nurse laughed and I collected myself and examination, a little bit later in another office that did have windows.

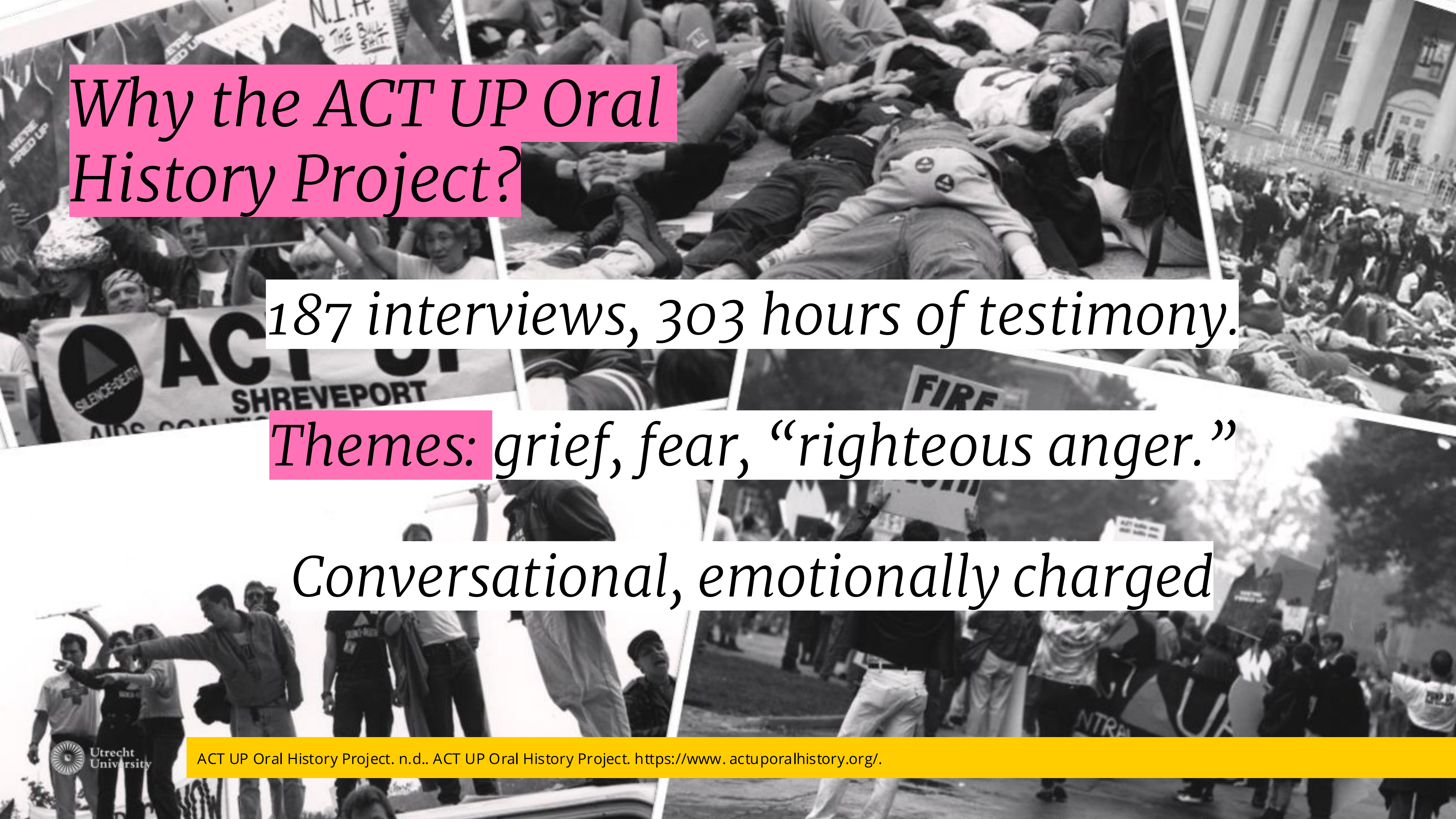
SS: So, then you had AIDS, as far as you were concerned.

MP: Yeah, as far as they were concerned, and I went into a huge thing of – you've made a big mistake, you've mixed up my biopsy with someone else's – denial, you know?

SS: So, what happened after that, in your life?

MP: I calmed down and N. Patrick Hennessey tells me, I've got about six months to a year, and that I have to very quickly go to the Gay Men's Health Crisis. I've got to interrupt myself at this point, because I remember a few years prior – one thing that

Utrecht University



Why the ACT UP Oral History Project?

187 interviews, 303 hours of testimony.

Themes: grief, fear, "righteous anger."

Conversational, emotionally charged

ACT UP Oral History Project. n.d.. ACT UP Oral History Project. https://www. actuporalhistory.org/.

Our approach – Overview

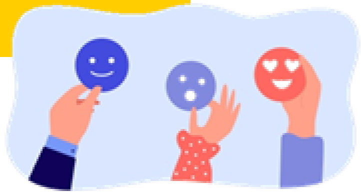
Audio

Subtitles

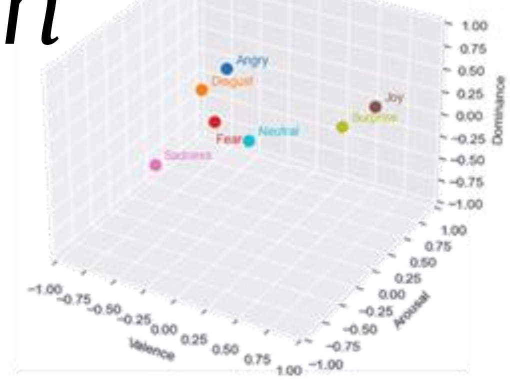
→ Automated Speech Preprocessing → Emotional Sentences Pre-Selection → Emotion Annotation Study → Automatic Emotion Recognition

↓

Predicted Emotion



Our approach – Sentence Pre-Selection



Audio

Subtitles

23 interviews (37 hours)

→ Automated Speech Preprocessing → Aligned and Diarized subtitles

For each sentence...

→ Text Emotion Prediction (DistilRoBERTa)

→ Speech Emotion Prediction (Speech Emotion Diarization)

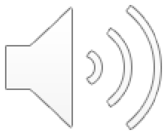
→ VAD Distance Calculation → Pre-Selected Emotional Sentences

728 sentences

Utrecht University

J. Hartmann. 2022. Fine-tuned RoBERTa models for emotion classification. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base Hugging Face.
Yingzhi Wang, Mirco Ravanelli, and Alya Yacoubi. 2023. Speech emotion diarization: Which emotion appears when?. In 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 1–7

Our approach – Sentence Pre-Selection

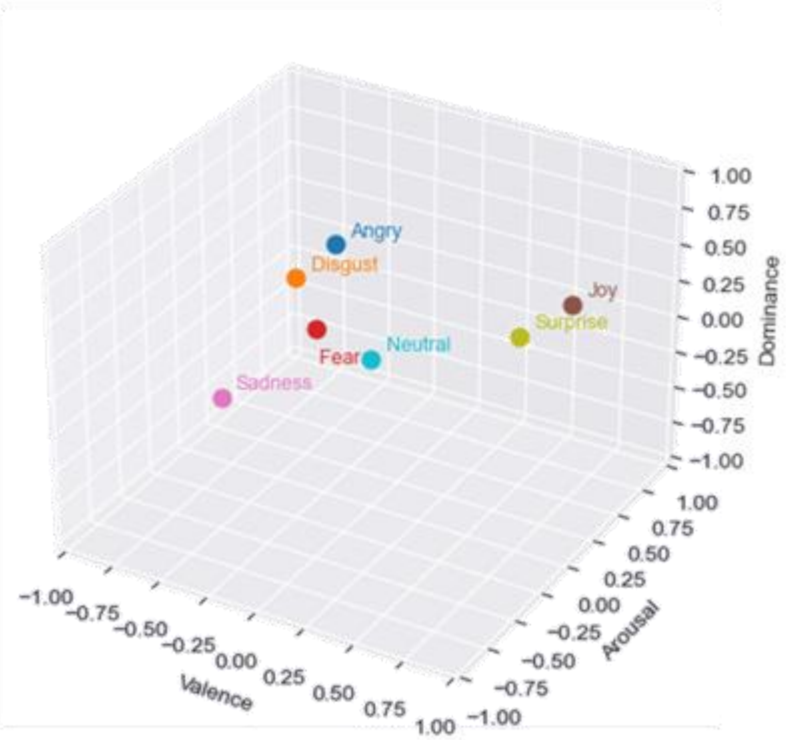


Neutral	Happy

5 sec 10 sec

Well, my grandmother gets all upset and says, you know, "Oh, this is some Mafia-controlled baby beauty contest!".

Anger



$$distance(paralinguistics, linguistics) = \frac{distance(neutral, anger) * t_{neutral} + distance(happy, anger) * t_{happy}}{t_{total}}$$

Our approach – Emotion Annotation Study (Trial 1)

“The whole art world, the trajectory of the art world, the marketplace, everything, it was frightening.”

→ Speech Emotion Annotation

→ Text Emotion Annotation

“Combination of anger and sadness but overall just blehhhh” (paralinguistics)

- ✓ Educational Setting
- ✓ 32 annotators
- ✓ 3 to 4 annotators per sentence (per modality)
- ✓ Annotations based on common sense

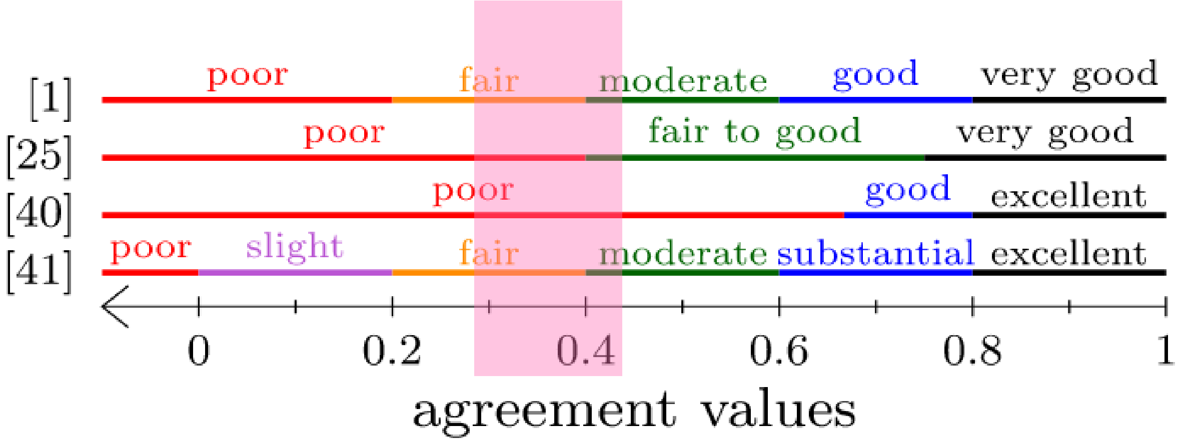
Emotions: 6 Ekman basic emotions (happy, sad, angry, fear, disgust, surprise) + neutral

Utrecht University

Our approach – Emotion Annotation Study (Trial 1)

Krippendorff's alpha:

- Linguistics: 0.42
- Paralinguistics: 0.31



Challenge #1: There are different interpretations for the Krippendorff's alpha making it challenging to interpret results consistently

Siegert, I., Böck, R. & Wendemuth, A. Inter-rater reliability for emotion annotation in human–computer interaction: comparison and methodological improvements. J Multimodal User Interfaces 8, 17–28 (2014). https://doi.org/10.1007/s12193-013-0129-9

Our approach – Emotion Annotation Study (Trial 1)

Let's consider that "good" corresponds to κ or α > 0.67.

J Multimodal User Interfaces (2014) 8:17–28
DOI 10.1007/s12193-013-0129-9

ORIGINAL PAPER

Inter-rater reliability for emotion ann interaction: comparison and methodo

Ingo Siegert · Ronald Böck ·

Table 6 Overview and the IRR for th 405

	Interaction type	Mu
Set 1	Random	Au
Set 2		Vid
Set 3		Au

Categories: sadness, helpless joy, surprise, confusion, ange emotion.

A Multimodal, Multilabel Approach to Recognize Emotions in Oral History Interviews

Anargh Viswanath*, Michael Gref†, Teena Hassan‡ and Christoph Schmidt§

Mess! ork for Building otated Corpus

lercq, Veronique Hoste
Ghent University
lgium
te}@ugent.be

Challenge 2: Subjective annotation tasks are difficult!

Emotions	Fleiss' Kappa (κ)↑
Happy	0.49
Sad	0.44
Anger	0.31
Surprise	0.10
Disgust	0.30
Fear	0.33

Emotions	HdG		CMU-MOSEAS	
	AUC	balAcc	AUC	balAcc
Happy	0.78	0.72	0.70	0.66
Sad	0.63	0.61	0.74	0.68
Anger	0.76	0.72	0.80	0.77
Surprise	0.81	0.75	0.60	0.59
Disgust	0.77	0.74	0.66	0.64
Fear	0.67	0.64	0.64	0.62
Overall*	**0.74**	**0.70**	**0.69**	**0.66**

*Mean average of all six emotions

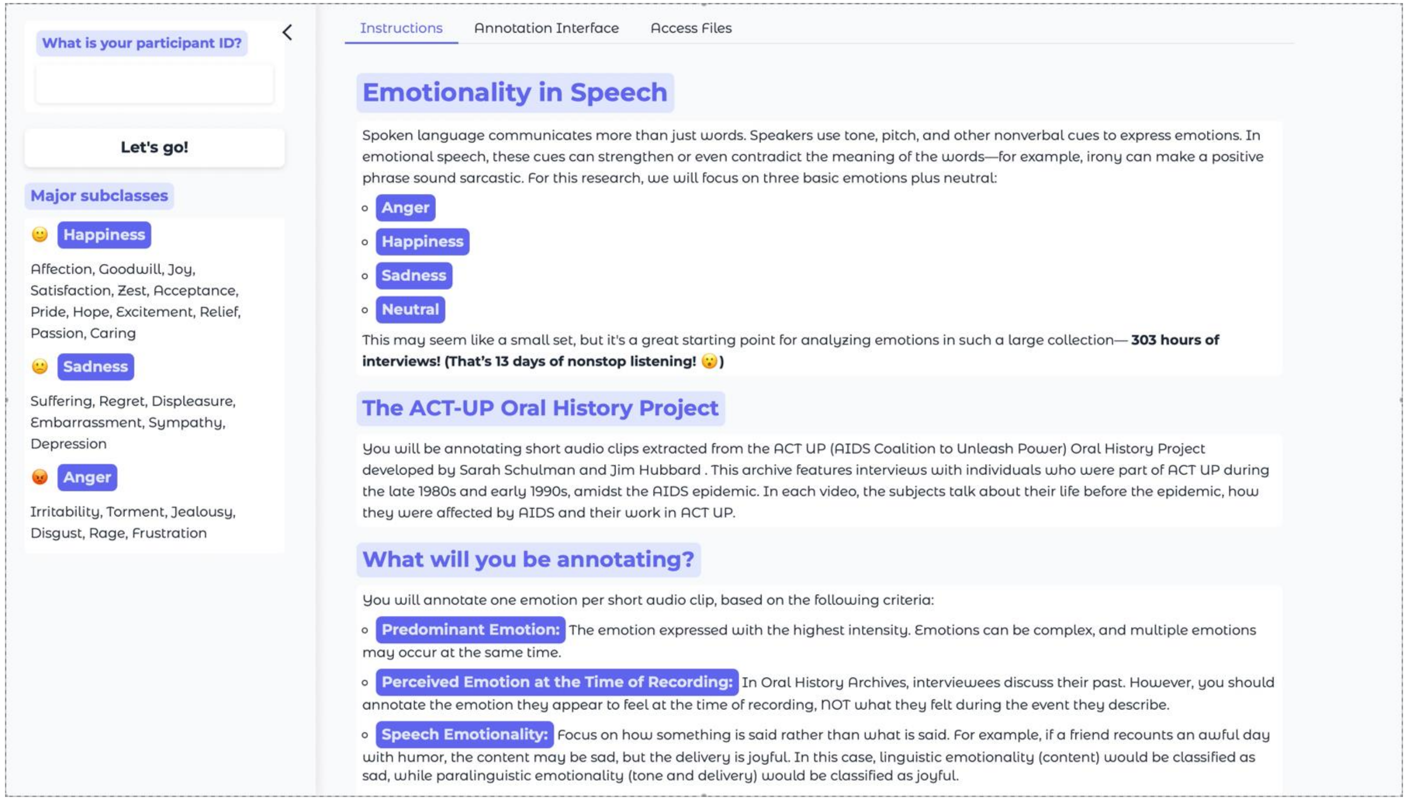
	κ
	1
ness	0,315
Optimism	0,438
Pity	0,597
Pride	0,524
Rejection	0,357
Relief	nan
Remorse	0,602
Sadness	0,678
Suffering	0,442
Surprise	0,079
Torment	0,01

emotion category.

Majority voted for model training

Utrecht University

Our approach – Emotion Annotation Study (Trial 2)



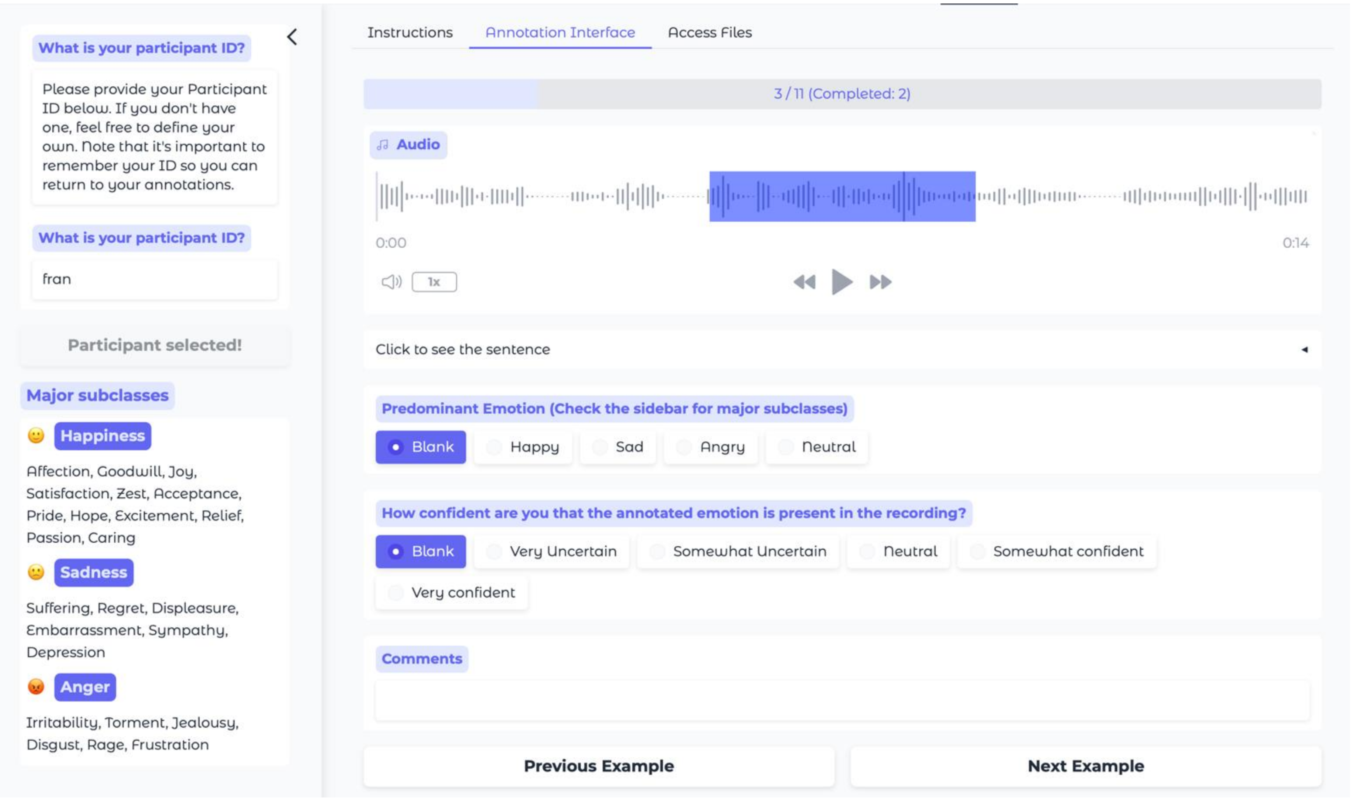
Open-source annotation tool developed for Trial 2

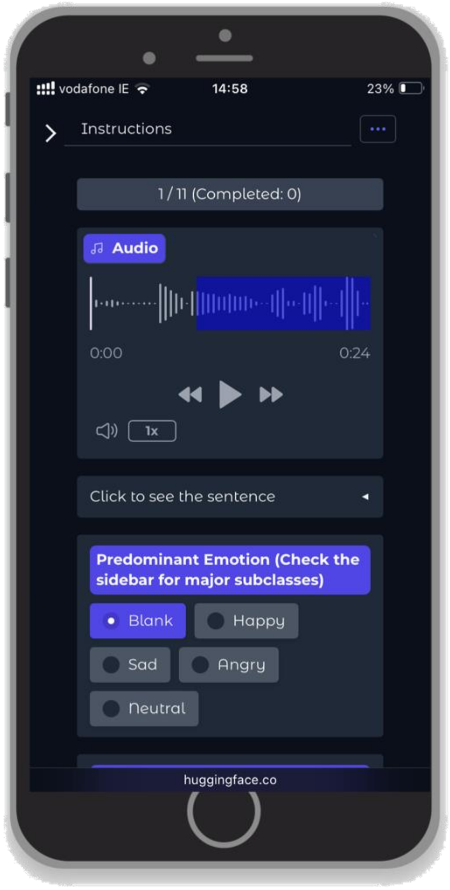
- ✓ Online Setting
- ✓ 18 annotators
- ✓ 2 annotators per sentence
- ✓ Context before and after the target sentence
- ✓ Extensive instructions

Emotions: happy, sad, angry and neutral

Utrecht University

Our approach – *Emotion Annotation Study (Trial 2)*





Try it out!

Utrecht University

Our approach – Emotion Annotation Study (Trial 2)

Participants' Annotations

Emotions

Happy, Sad, Angry, Neutral

Confidence

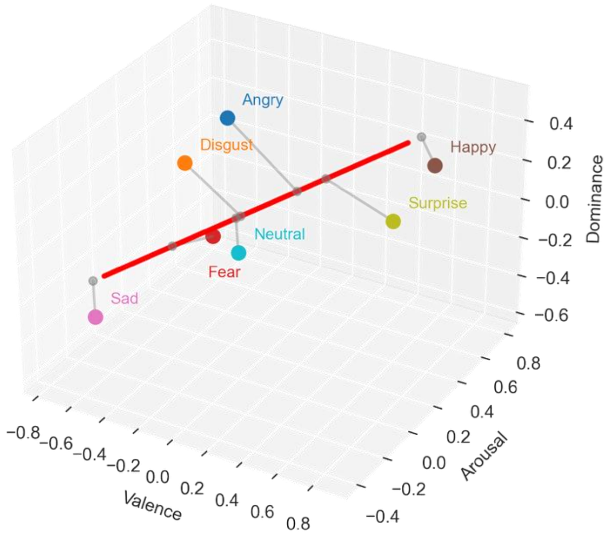
Very uncertain → Very confident

→

Nominal Krippendorff's α

Proposed Weighted (Ordinal) Krippendorff's α

Correlation between confidence and agreement



Visualization of emotion coordinates in the VAD space with the 1st PCA direction.

Our approach –Emotion Recognition

Audio

Text

Participants' Annotations

728 sentences

2 annotators per sample

Paralinguistic Feature Extraction
(Wav2Vec2-based models)

Linguistic Feature Extraction
(RoBERTa-based models)

4-Class Emotion Recognition Classifier (SVM)

Emotion-Specific Regression Models (SVR)

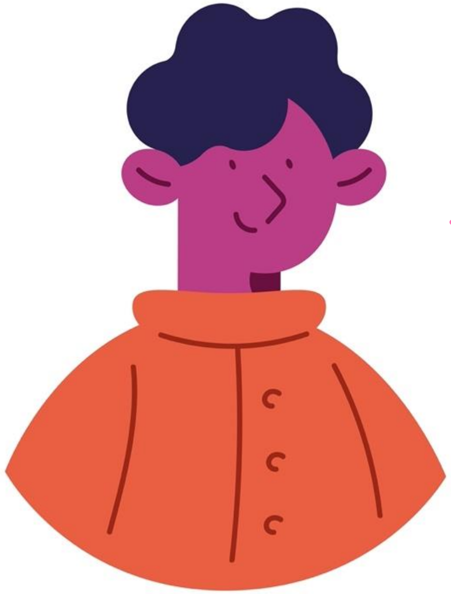
All models were trained on text, speech and both combined

Emotion
(1 emotion per sample)

Emotion
(up to 4 emotions per sample)

Our approach – Emotion Recognition

"And I felt scared, you know, because even inside the city hall, they're gunning us down."



I am very confident sadness is present in this example.

Annotator 1

I am somewhat uncertain happiness is present in this example.

"The high pitch in the word "scared" makes me feel like he's happy saying it, which is why I used "happy" — it sounds like he's relieved while saying it."

Annotator 2

Utrecht University

Our approach – Emotion Recognition

4-Class Emotion Recognition Classifier

Sample 1.1

Audio

Sad

Sample 1.2

Audio

Happy

Each annotation is seen as an independent sample for both training and testing.

4 Emotion-Specific Regression Models

Sample 1 (Happy, Sad, Angry, Neutral)

(1, 1, 0, 0)
or
weighted for confidence
(2, 5, 0, 0)

Annotations are combined into a single ground truth label.
The threshold for emotion binarization defined iteratively during cross-validation.

Insights – *Emotion Annotation Study*

Participants' Annotations	*Nominal α*	*Nominal α (Trial 2) ≈ 0.30* *Ordinal α (Trial 2) ≈ 0.40*
Emotions *Happy, Sad, Angry, Neutral* →	*Ordinal α*	*Trial 2 has a lower stardard deviation for α, suggesting more consistent agreement across annotator pairs.*
Confidence *Very uncertain → Very confident*	*confidence vs agreement*	*Confidence positively correlated with agreement ($r = 0.20$, $p < 1e{-}14$)*

Insights –Emotion Recognition

4-Class Emotion Recognition Classifier → *Emotion (1 emotion per sample)*

Emotion-Specific Regression Models → *Emotion (up to 4 emotions per sample)*

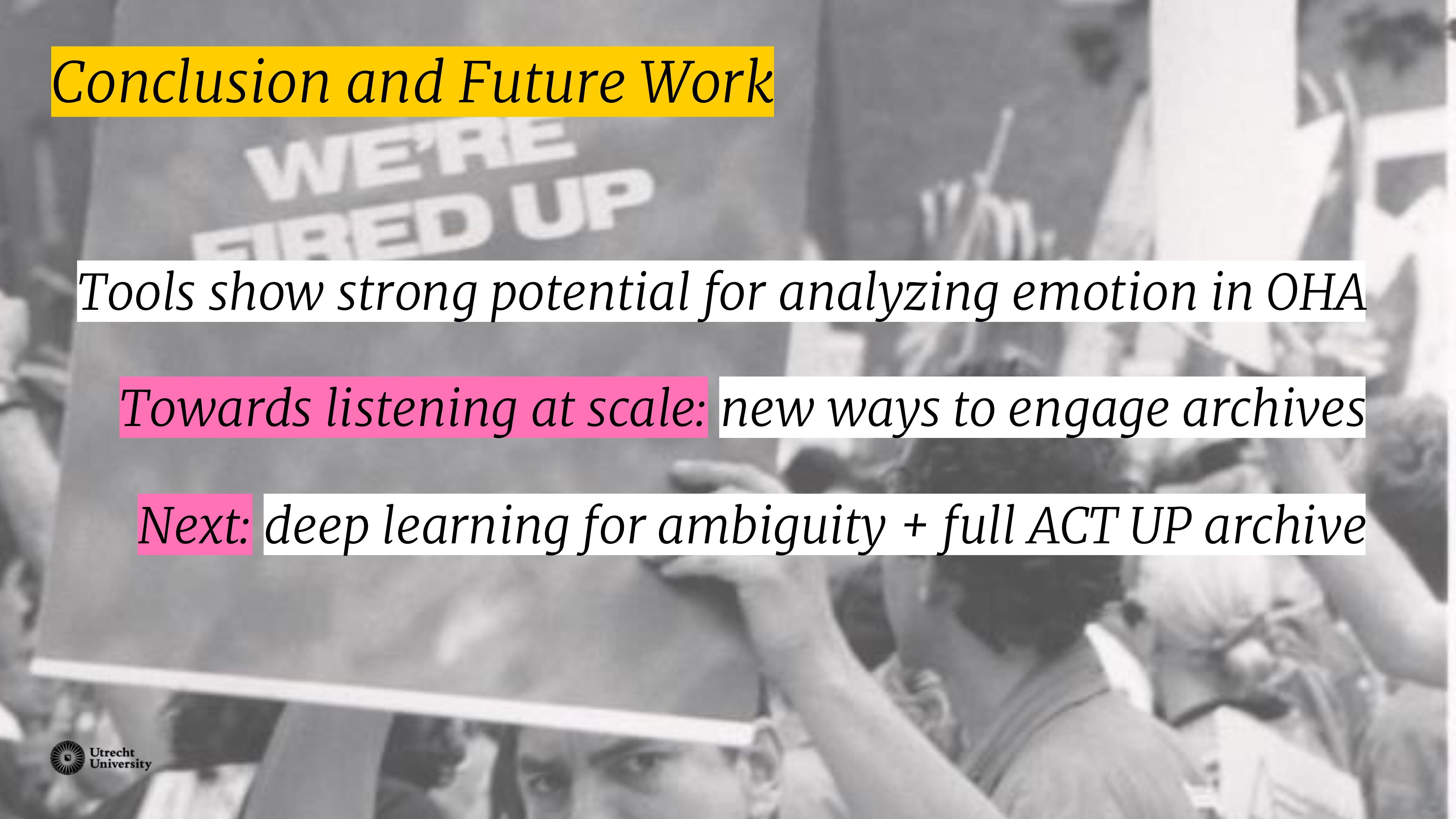
All models were trained on text, speech and both combined

Speech > Text for emotion recognition

Best model: F1 ≈ 0.66 (emotion-specific regression, paralinguistic modality)

Comparable to state-of-the-art for emotion prediction in Oral History Archives

Utrecht University



Conclusion and Future Work

Tools show strong potential for analyzing emotion in OHA

Towards listening at scale: new ways to engage archives

Next: deep learning for ambiguity + full ACT UP archive

Utrecht University

Sharing science,
shaping tomorrow

Contact:
f.pessanha@uu.nl

GitHub:
github.com/franciscapessanha/emotion-annotation-with-gradio

Icons designed by Freepik
Archival images from the Branson Collection, History at NIH