AI-BASED VISUALIZATION OF VOICE CHARACTERISTICS IN LECTURE VIDEOS' CAPTIONS
AGENDA

Introduction

Related Work

AI-based Visualization of Voice Characteristics in Captions

User Study

Conclusion and Future Work
INTRODUCTION
100+ STUDIES:
CAPTIONING a video improves
+ COMPREHENSION of
+ ATTENTION to
+ and MEMORY for
the video

So even though we face the difficulties
100+ STUDIES:
CAPTIONING a video improves
+ COMPREHENSION of
+ ATTENTION to
+ and MEMORY for
the video

Contribution to making messages clear in human communication:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WORDS</td>
<td>7%</td>
</tr>
<tr>
<td>TONE</td>
<td>38%</td>
</tr>
<tr>
<td>INTONATION</td>
<td></td>
</tr>
<tr>
<td>VERBAL PACE</td>
<td></td>
</tr>
</tbody>
</table>

So even though we face the difficulties
NO PROGRESS IN CAPTIONS & SUBTITLES

for decades

So even though we face the difficulties

limited boring

No information from the voice
No emotions from the voice

AI-based Visualization of Voice Characteristics in Lecture Videos' Captions. Tim Schlippe, Katrin Fritsche, Ying Sun, and Matthias Wölfel. AIET 2022.
AI-based Visualization of Voice Characteristics in Lecture Videos’ Captions. Tim Schlippe, Katrin Fritsche, Ying Sun, and Matthias Wölfel. AIET 2022.
I have a dream that one day
Nooo

There are some Gauls on the building site. Three of them.

What? An old druid with a white beard, a cunning little one and a stupid great oaf!

AARGH!

The name, Iron Man, is... Thanos!

More properly, Thanos the first...

--Emperor shortly of Near-Defeated Titan-- Then of your own Earth!

Don't stick out your boots for me to lick, Smiley!

Weak or Fading Speech

Oooo-- HHH!!

Eyeshield 21, 2006

Asterix – The Gaul, 1961

Iron Man, 1973
RELATED WORK
RELATED WORK:
Font Development and Representation

Typography as stylish device

Dadaismus 1917
Asterix – The Gaul 1961

11 Ai-based Visualization of Voice Characteristics in Lecture Videos' Captions. Tim Schlippe, Katrin Fritsche, Ying Sun, and Matthias Wölfel. AIET 2022.
RELATED WORK: Font Development and Representation

“Digital reading media [has to be treated] as a system of variable and dynamic design elements for text presentation and text accessibility in communicative spheres.” (Kuhn, 2017)

Typography as stylish device

Responsive Type

(Lee et al., 2006)
(Rashid et al., 2008)

(Wölfel and Stitz, 2015)

Dadaismus  Asterix – The Gaul

1917  1961  2000
RELATED WORK:
Font Development and Representation

“Digital reading media [has to be treated] as a system of variable and dynamic design elements for text presentation and text accessibility in communicative spheres.” (Kuhn, 2017)

**Typography as stylish device**

**Responsive Type**

(Lee et al., 2006)
(Rashid et al., 2008)

(Wölfel and Stitz, 2015)

**Voice Driven Type Design**

(Wölfel et al., 2015), (Schlippe et al, 2018),
(Bessemans et al., 2019)
(de Lacerda Pataca and Costa, 2020)
(de Lacerda & Dornhofer Paro Costa, 2022)
RELATED WORK:
Captioning and Subtitling

Computer assisted captioning

(Martone et al., 2004)
(Boulianne et al., 2006)
(Levin et al., 2014)
(Tremblay et al., 2015)
RELATED WORK: Captioning and Subtitling

Computer assisted captioning
(Martone et al., 2004)
(Boulianne et al., 2006)
(Levin et al., 2014)
(Tremblay et al., 2015)

Placement and design
(Vy and Fels, 2010)
(Brown et al., 2015)
(Fox, 2016)
RELATED WORK: Captioning and Subtitling

Computer assisted captioning
(Martone et al., 2004)
(Boulianne et al., 2006)
(Levin et al., 2014)
(Tremblay et al., 2015)

Placement and design
(Vy and Fels, 2010)
(Brown et al., 2015)
(Fox, 2016)

Emojis and emoticons
(El Taweel, 2016)

2000 2010 2016
RELATED WORK: Captioning and Subtitling

Computer assisted captioning
(Martone et al., 2004) (Boulianne et al., 2006) (Levin et al., 2014) (Tremblay et al., 2015)

Placement and design
(Vy and Fels, 2010) (Brown et al., 2015) (Fox, 2016)

Emojis and emoticons
(El Taweel, 2016)

WaveFont
(Schlippe et al., 2020) (Schlippe et al., 2022)
AI-BASED VISUALIZATION OF VOICE CHARACTERISTICS IN CAPTIONS
TECHNICAL IMPLEMENTATION:
Visualization of Voice Characteristics

1. Extraction of the audio track from the video

00:02.560 --> 00:04.350 So even though
00:04.864 --> 00:05.384 we face ...
TECHNICAL IMPLEMENTATION:
Visualization of Voice Characteristics

2. Segmentation into smaller audio files
3. Automatic forced alignment
4. Acoustic feature extraction

So | even | though

quiet & fast | loud & slow | loud & fast
5. Mapping of acoustic features to font classes

- So: thin & narrow font
- even: bold & wide font
- though: bold & narrow font
TECHNICAL IMPLEMENTATION: Visualization of Voice Characteristics

6. Type design
WAVEFONT IN LECTURE VIDEOS’ CAPTIONS

unique & innovative

We make emotions readable.

STUDY

✓ visualization of voice characteristics, ✓ understanding the content, ✓ linguistic understanding, ✓ following the content, ✓ identifying important words
WAVEFONT IN LECTURE VIDEOS’ CAPTIONS

Visualization of Voice Characteristics

Preferences in terms of captions in lecture videos (1st, 2nd, 3rd priority).

Assumption of improved WaveFont comprehension with more practice.
CONCLUSION & FUTURE WORK
CONCLUSION AND FUTURE WORK

Conclusion

– AI-based visualization of voice characteristics in captions helps students improve the viewing and learning experience in lecture videos.

– *WaveFont* technology processes the speech signal and intuitively displays loudness, speed, and pauses in the subtitle font.

– In our survey, the AI-based visualization of speech features outperformed standard captions.

– Participants agree that they assume to understand *WaveFont* with more practice even better.

Future Work

– Analyze learning effect in more detail.

– Investigate effect of language acquisition for non-native speakers.

– Analyze the effect of *WaveFont* captions on learning styles.
THANK YOU