Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic

Tim Schlippe, Shaimaa Alessai, Ghanimeh El-Taweel,
Matthias Wölfel and Wajdi Zaghouani

The 19th International Conference on Cyberworlds (CW 2020)
29 September – 1 October 2020

Captions & subtitles
are in high demand

466 millions
hearing-impaired
(WHO, 2020)

30 millions
watch with captions
(Amazon, 2016)

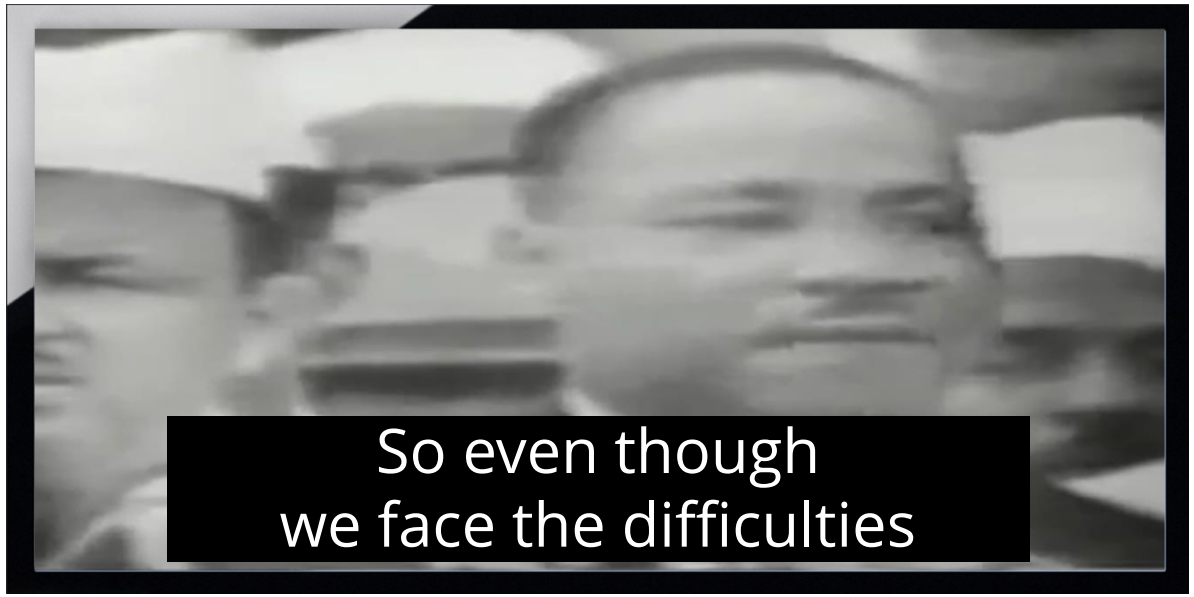
244 millions
migrants
(UN, 2016)

2 billions
watch w/o sound
(Digiday, 2016)

Challenge

Captions & subtitles have not made progress

for decades



limited

boring

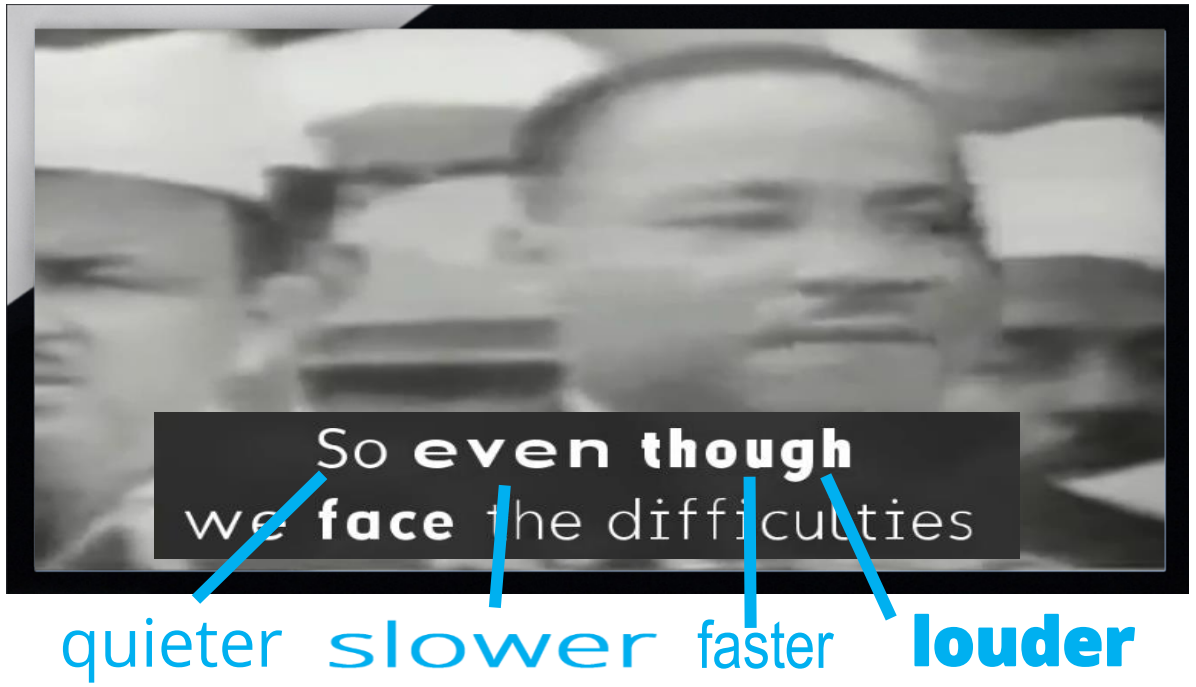
No information from the voice

No emotions from the voice

Solution: WaveFont captions

Extra-feature for captions & subtitles

unique & innovative



informative

Information from the voice

emotional

Emotions from the voice

Experimental setup
Visualization of voice characteristics

Experimental setup
Visualization of voice characteristics

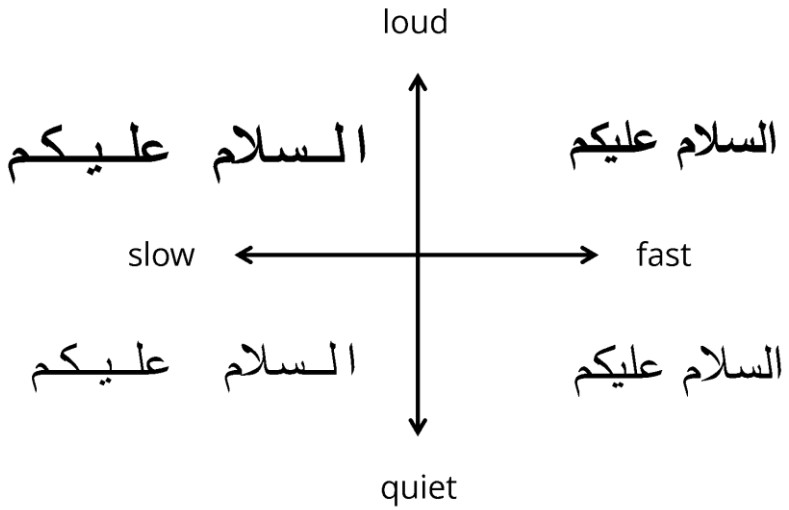
Simplified Arabic

DINArabic

Experimental setup
Visualization of voice characteristics

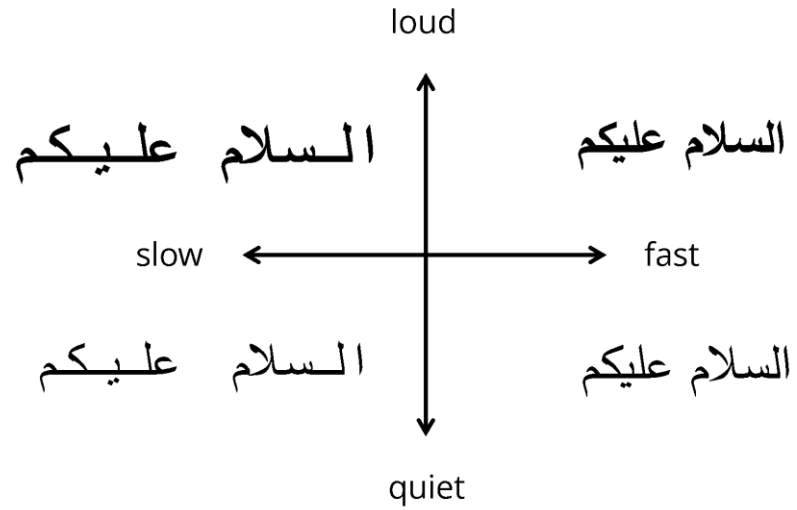
Simplified Arabic

DINArabic

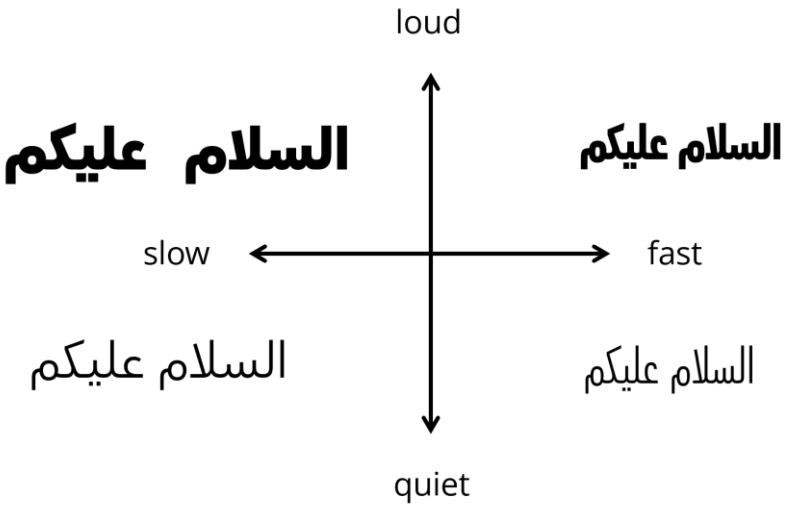


Experimental setup
Visualization of voice characteristics

Simplified Arabic



DINArabic

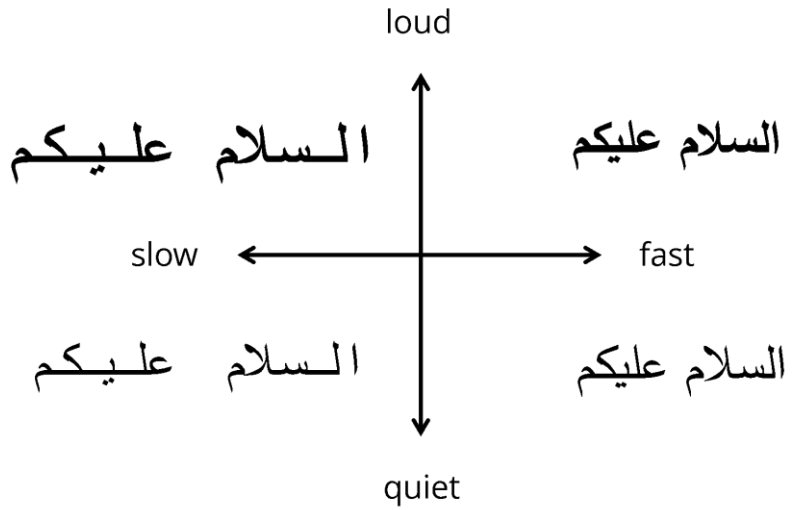


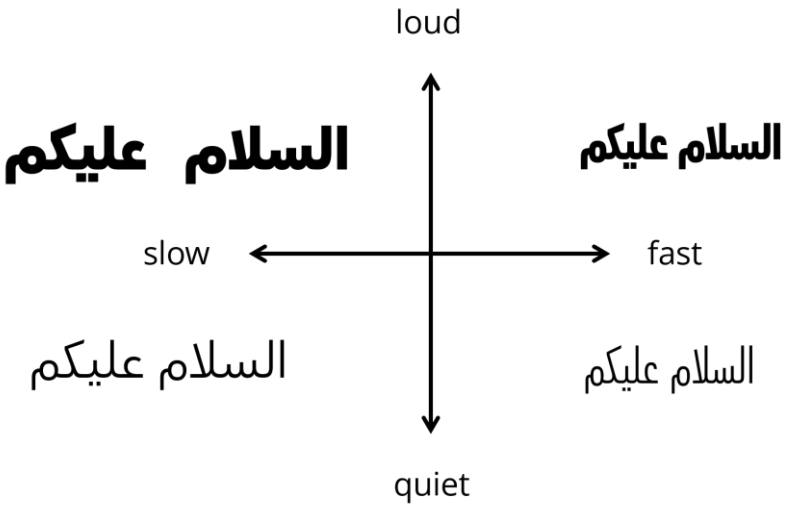
Experimental setup
Videos

76 seconds

Simplified Arabic

DINArabic



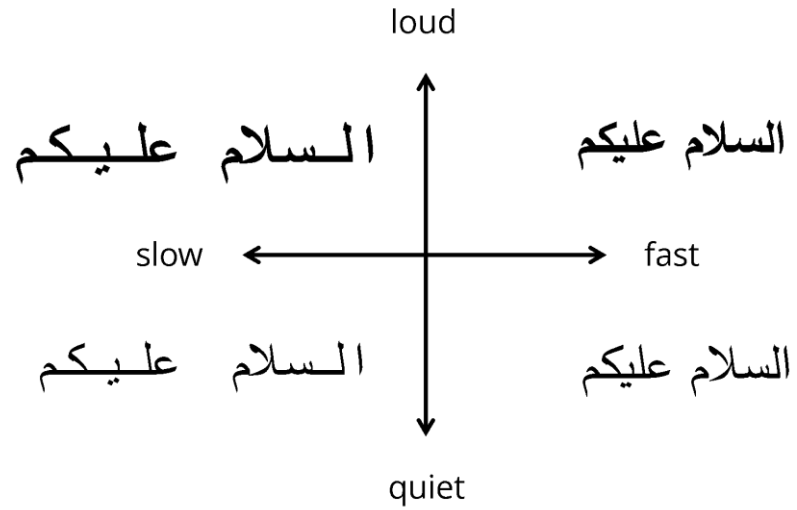


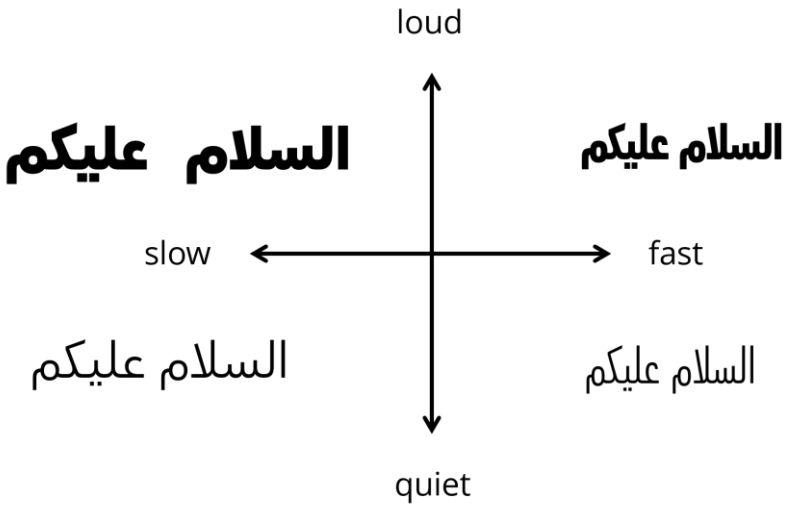
Experimental setup
Videos

76 seconds

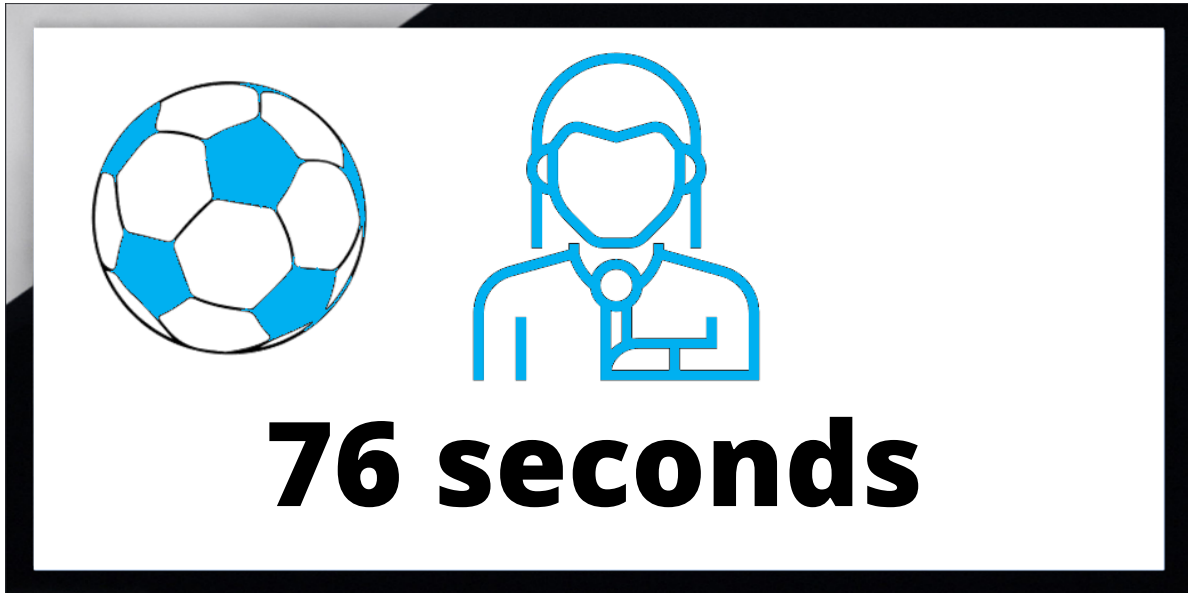
Simplified Arabic

DINArabic

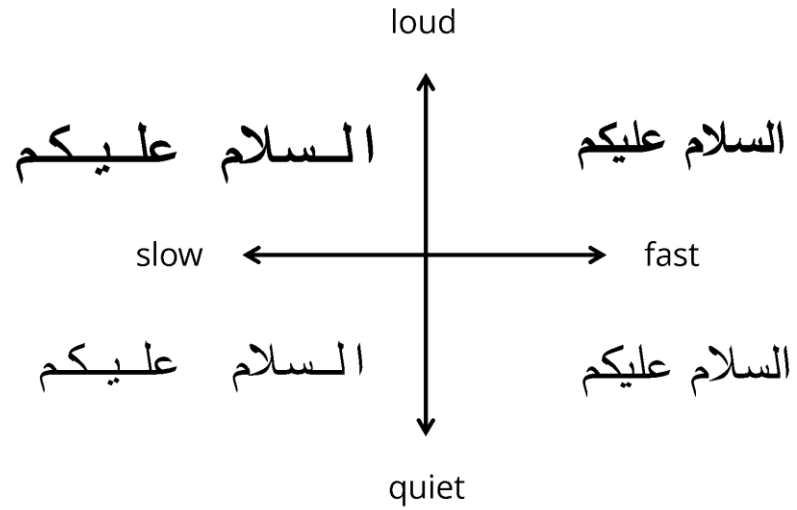




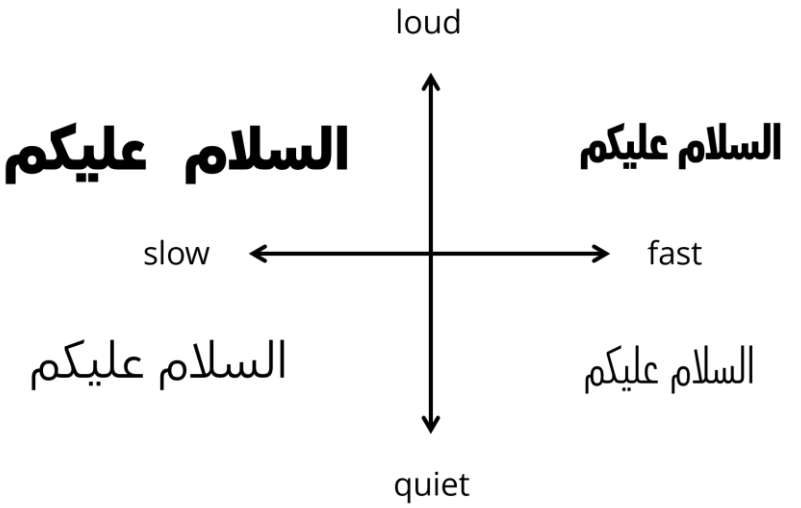
Experimental setup
Videos



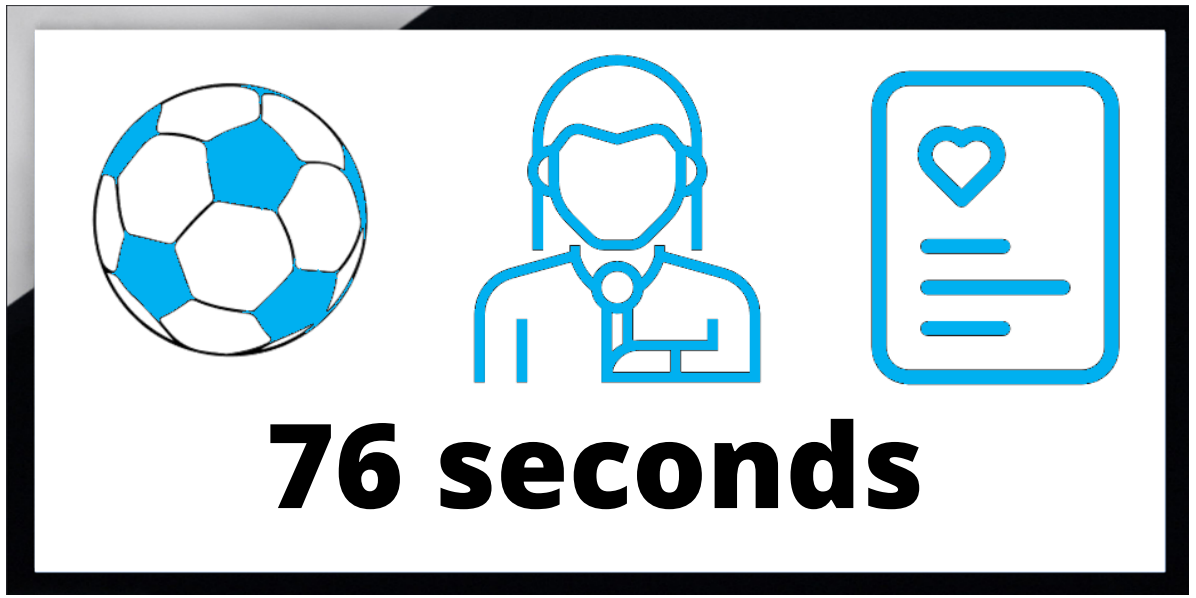
Simplified Arabic



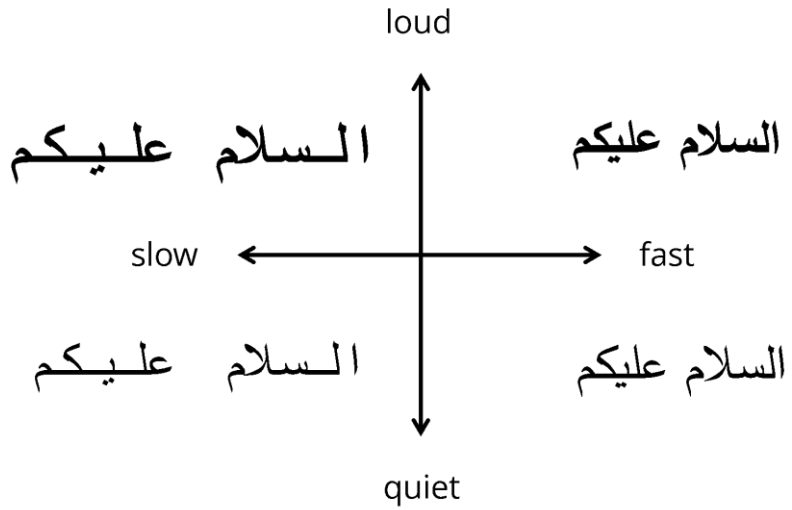
DINArabic



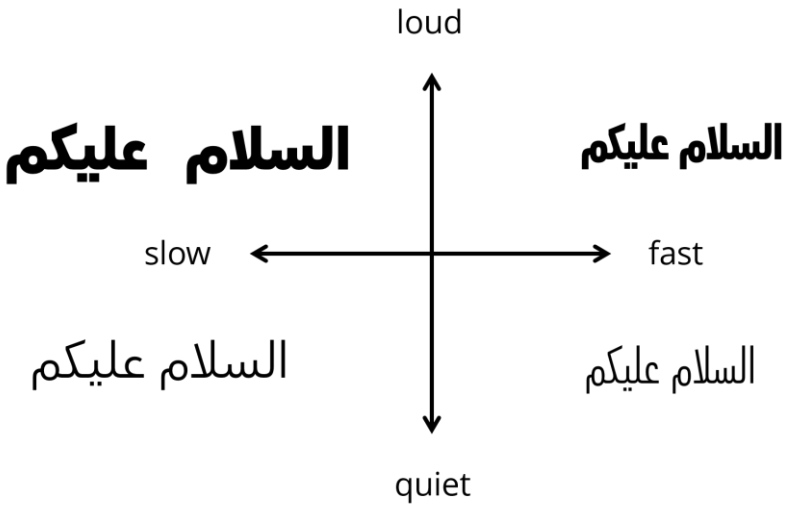
Experimental setup
Videos



Simplified Arabic



DINArabic



Experimental setup
Questionnaire with video clips

Brief introduction
and
general question about
the *WaveFont*
visualization

Personal questions
(about general subtitle
usage and demographic
information)

Video with
conventional
standard captions

sports, news, poetry

+

related questions

Video with *WaveFont*
captions visualized
with the
Simplified Arabic font

sports, news, poetry

+

related questions

Video with *WaveFont*
captions visualized
with the
DINArabic font

sports, news, poetry

+

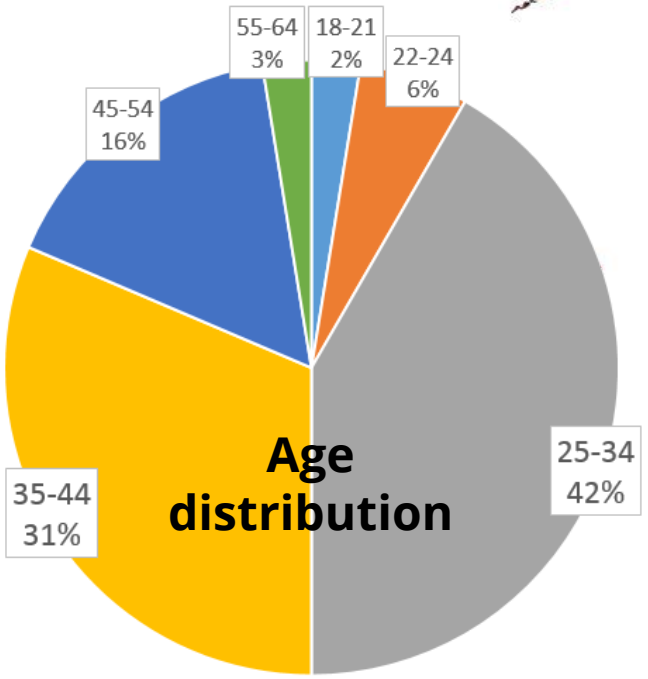
related questions

Experimental setup
Participants

79 x male

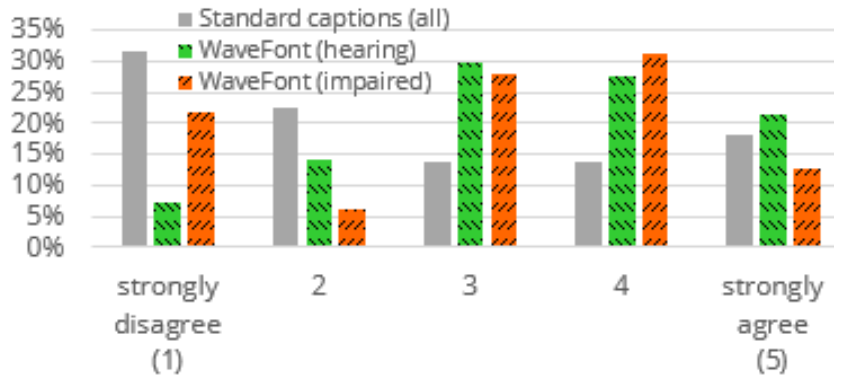
80 x female

32 x



Results

Visualization of voice characteristics comprehensible

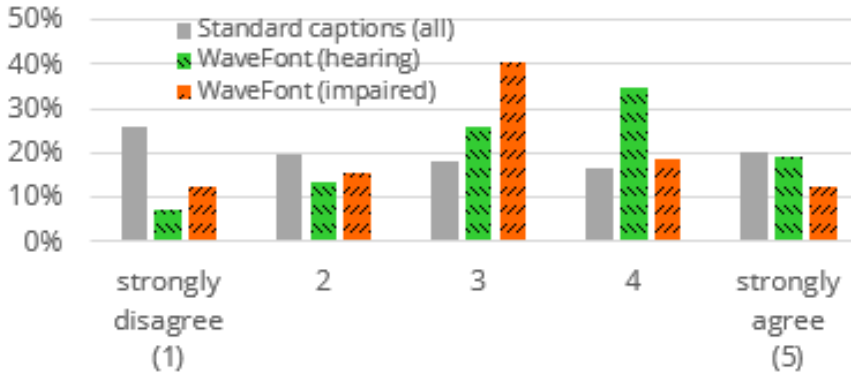


ø 2.6

ø 3.4

ø 3.1

Loudness comprehensible?



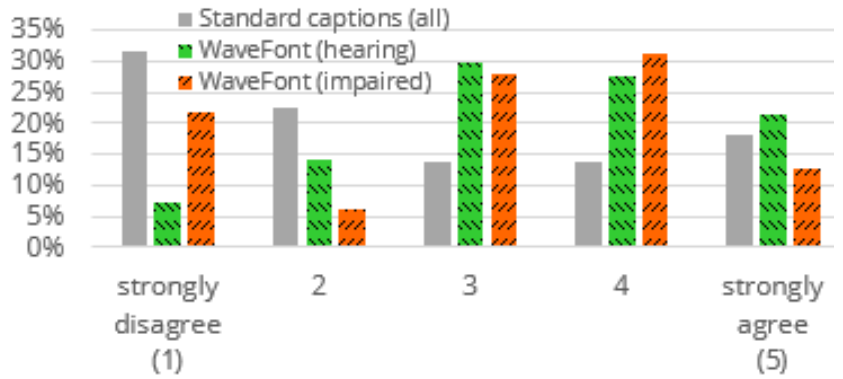
ø 2.6

ø 3.4

ø 3.0

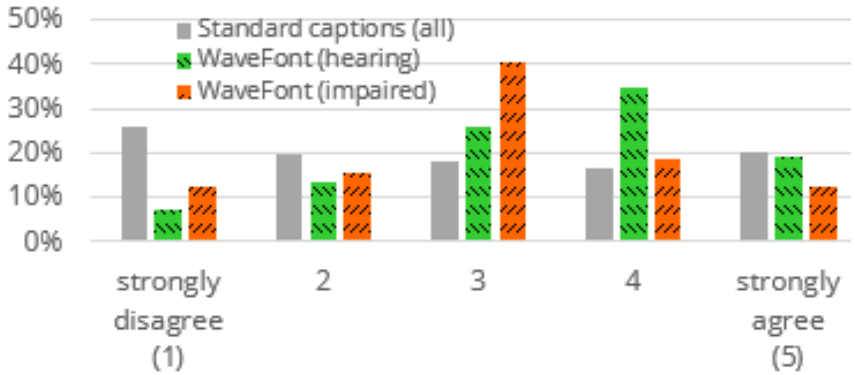
Speed comprehensible?

Results
Visualization of voice characteristics comprehensible



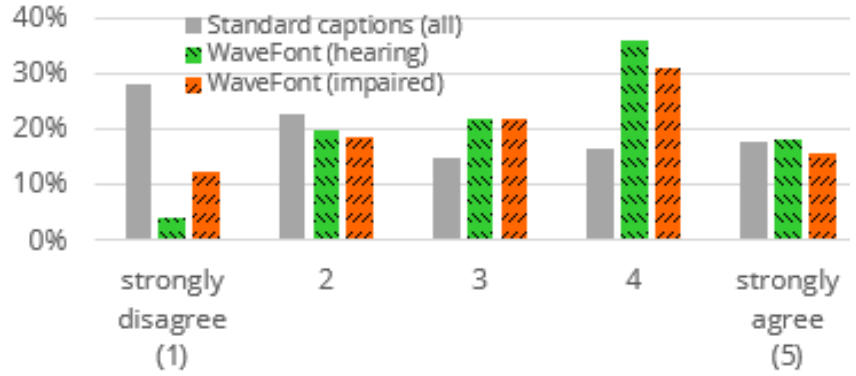
ø 2.6
ø 3.4
ø 3.1

Loudness comprehensible?



ø 2.6
ø 3.4
ø 3.0

Speed comprehensible?

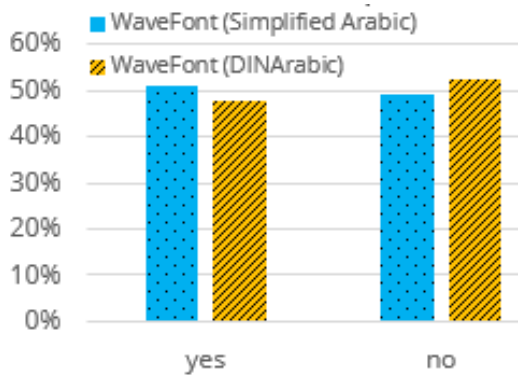


ø 2.8
ø 3.6
ø 3.2

Loudness and **speed** together comprehensible?

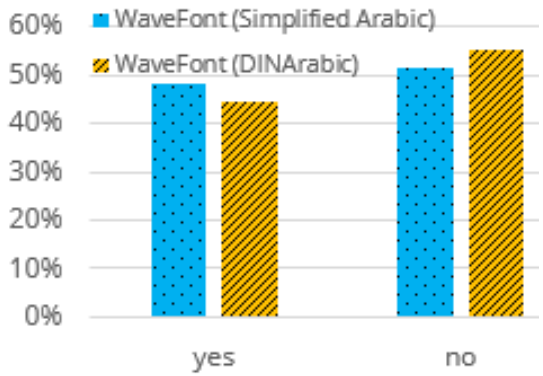
- *WaveFont* more comprehensible than standard
- Loudness and speed together are preferred
- Hearing rate slightly better than hearing-impaired

Results
Finer visualization of voice characteristics



52% 48% 48% 52%

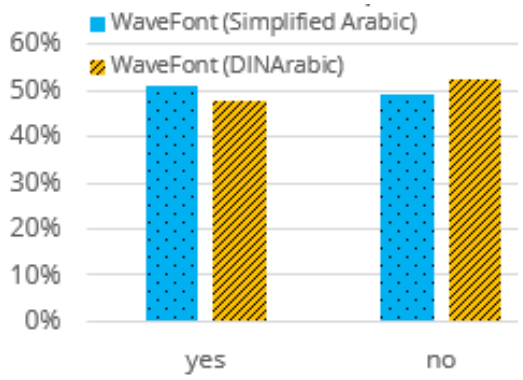
Third visualization to distinguish between **loud**, **normal** and **quiet**?



46% 54% 44% 56%

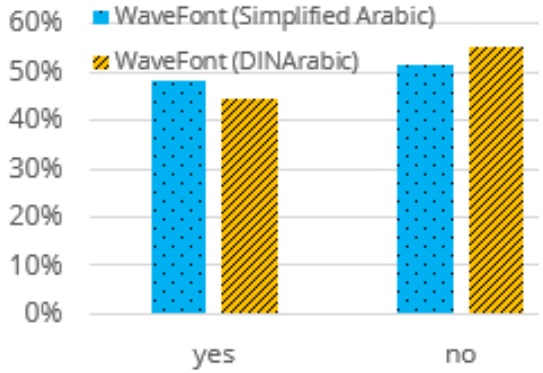
Third visualization to distinguish between **fast**, **normal** and **slow**?

Results
Finer visualization of voice characteristics



52% 48% 48% 52%

Third visualization to distinguish between **loud**, **normal** and **quiet**?

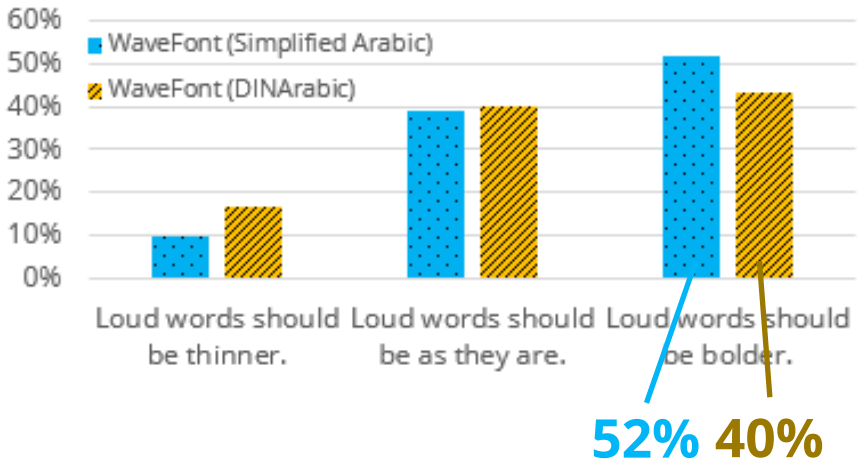


46% 54% 44% 56%

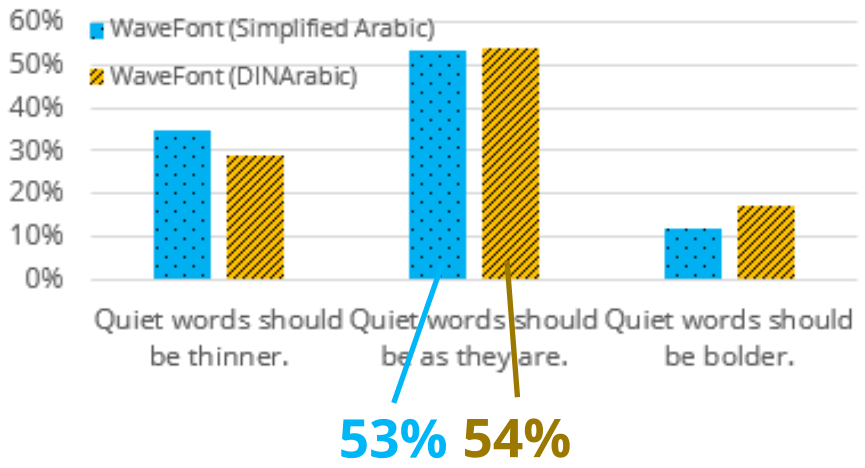
Third visualization to distinguish between **fast**, **normal** and **slow**?

- No clear majority for a third visualization at volume
- Two classes are preferred for speed

Results
Font selection



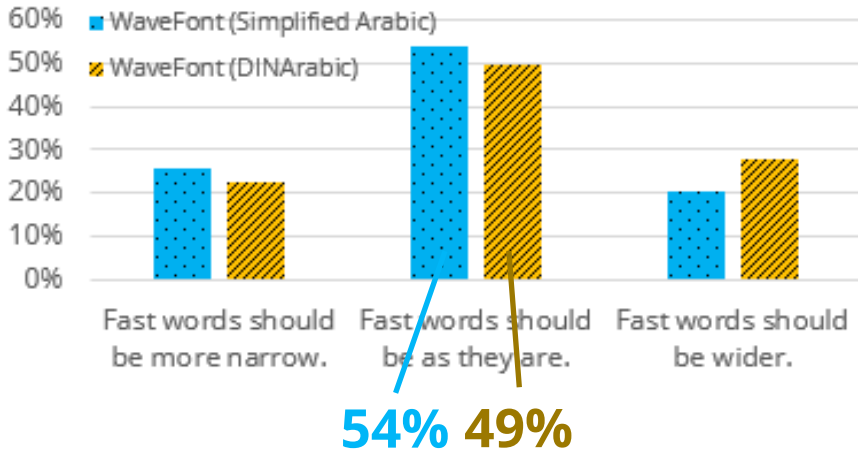
How to display **loud** words?



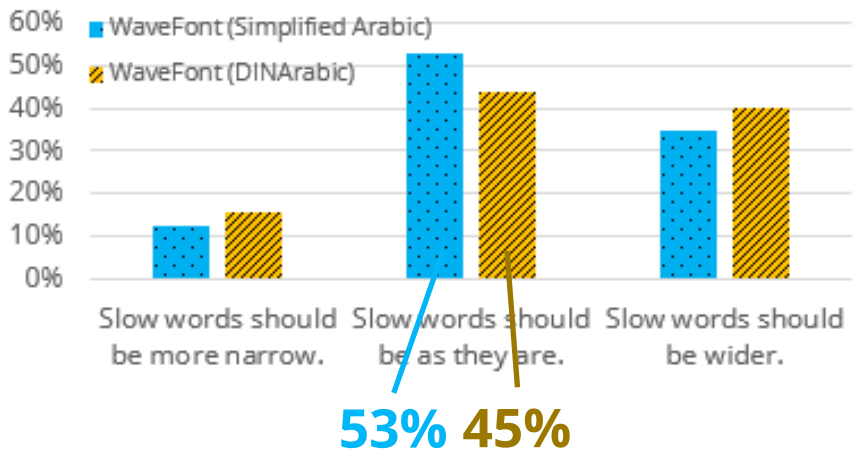
How to display **quiet** words?

- Loud words could be displayed even bolder.
- The majority is satisfied with the stoke width of quiet words.

Results
Font selection



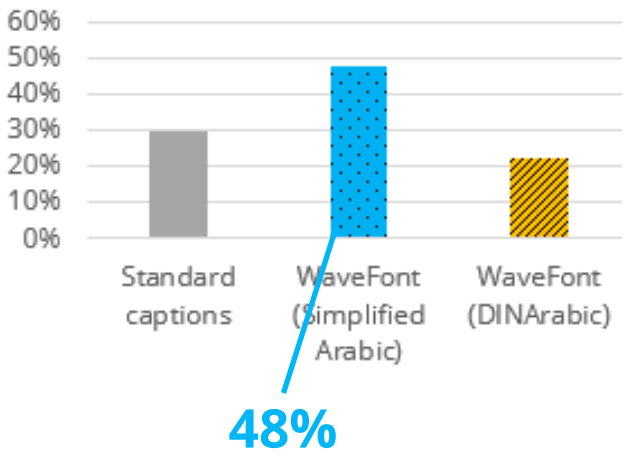
How to display **fast** words?



How to display **slow** words?

- The majority is satisfied with the character width of fast words.
- The majority is satisfied with the character width of slow words.

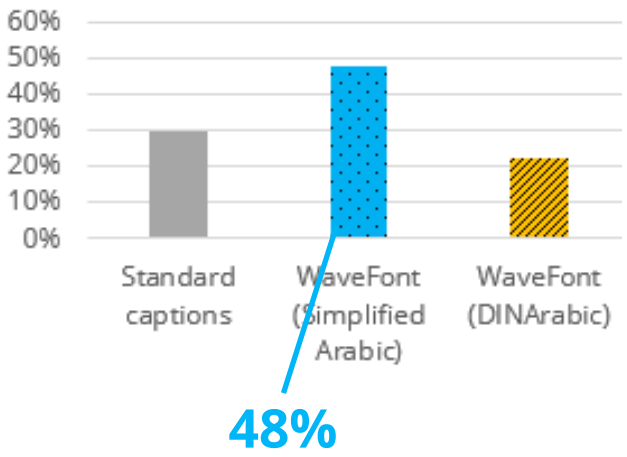
Results
Font selection



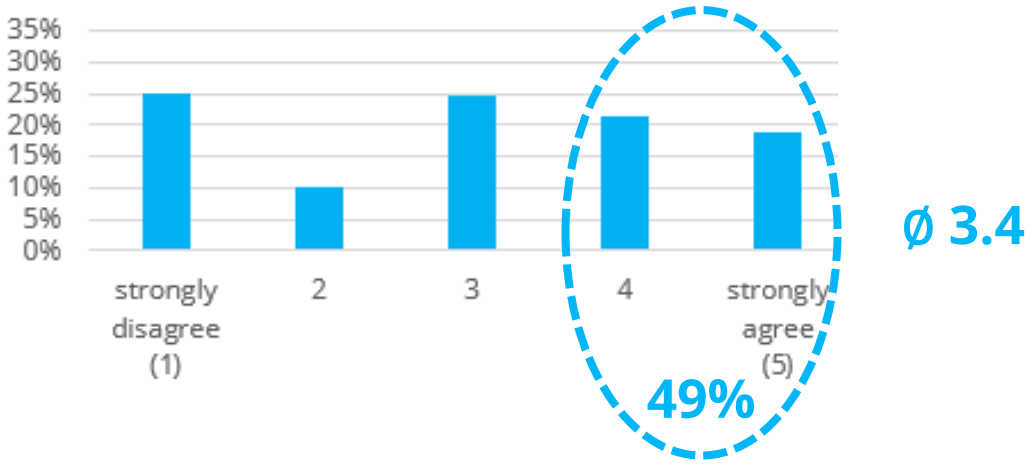
Which captions improve your TV/cinema experience more?

➡ Simplified Arabic – the more moderate font – is preferred.

Results
Font selection



Which captions improve your TV/cinema experience more?



Prefer movie that optionally displays ***WaveFont* captions**?

- Simplified Arabic – the more moderate font – is preferred.
- Movie with optional *WaveFont* captions is preferred.

Future work
Further experiments with Arabic

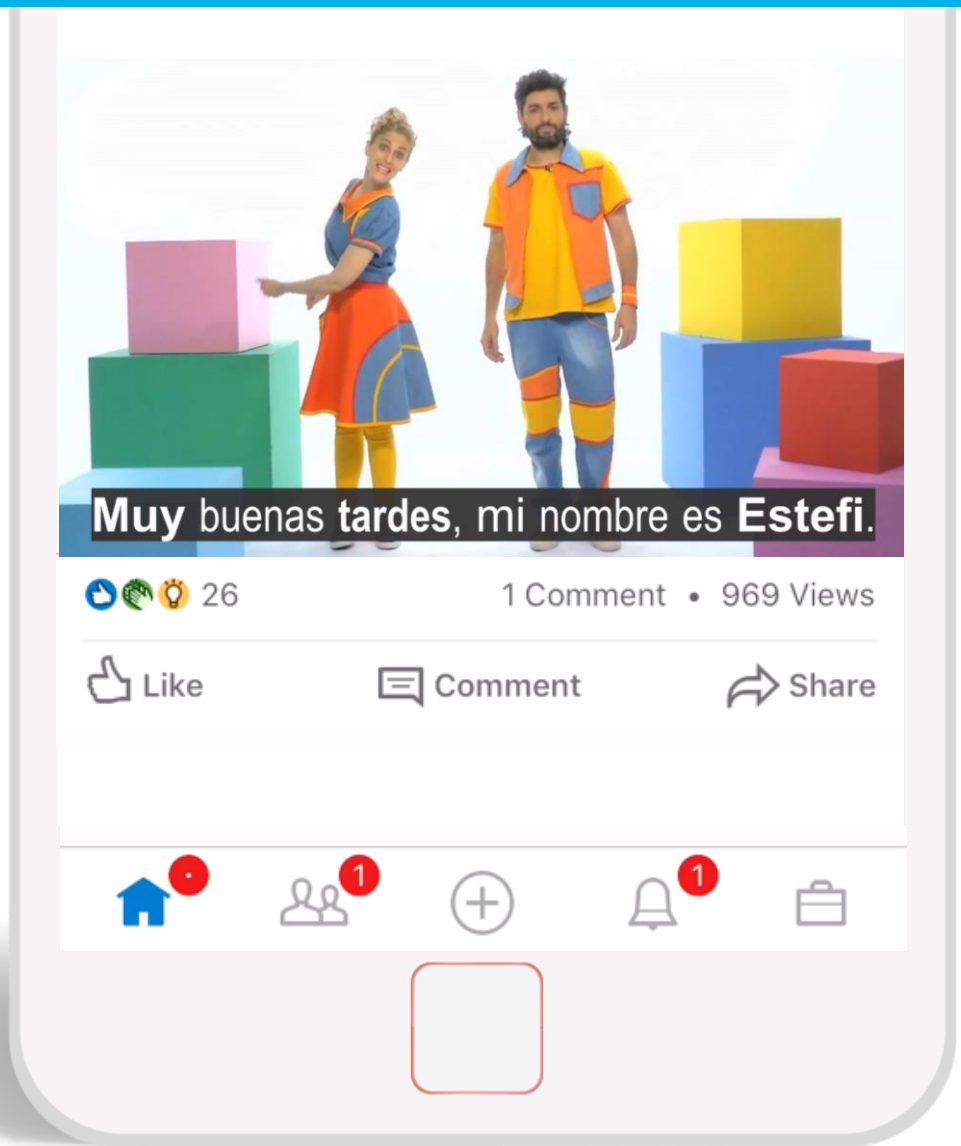
Mapping of the speech in Arabic dialect to words in the MSA subtitle

Investigate alternative visualizations of long words

- connective characters with the kashida (-)
- automatically inserting long vowels
- automatically inserting the diacritics

Future work

Investigate further languages and writing systems



Thank you

Research: t.schlippe@iubh.de

Business: tim.schlippe@silicon-surfer.com

References

[1] J. Ohene-Djan, J. Wright, and K. Combie-Smith, "Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People," in *CVHI*, 2007.

[2] M. Wölfel, T. Schlippe, and A. Stitz, "Voice Driven Type Design," in *SpeD*, 2015.

[3] "Arabic Alphabet," *Encyclopaedia Britannica*, Retrieved 2020-05-29.

[4] P. Shaw, "Codex: The Journal of Letterforms," *The Menhart Issue. John Boardley*, 2012.

[5] A. Kuhn and S. Hagenhoff, "Kommunikative statt objektzentrierte Gestaltung: Zur Notwendigkeit veränderter Lesekonzepte und Leseforschung für digitale Lesemedien," *Lesen X. 0. Rezeptionsprozesse in der digitalen Gegenwart. Göttingen*, pp. 27–45, 2017.

[6] M. Wölfel, "Rethinking text: Unleashing the full potential of media to provide a better reading experience," *IJACDT*, vol. 7, no. 2, pp. 1–11, 2018.

[7] J. Lee, S. Jun, J. Forlizzi, and S.E. Hudson, "Using Kinetic Typography to Convey Emotion in Text-Based Interpersonal Communication," in *DIS*. 2006, Association for Computing Machinery.

[8] R. Rashid, Q.V. Vy, R.G. Hunt, and D.I. Fels, "Dancing with Words: Using Animated Text for Captioning," *International Journal of Human-Computer Interaction*, vol. 24, pp. 505–519, 2008.

[9] M. Wölfel and A. Stitz, "Responsive Type - Introducing Self-Adjusting Graphic Characters," in *Interspeech*, 2015.

[10] M. Wölfel and D. Hill, "Acoustic Reading Experience: Aligning Sound Events to Text Using Gaze Tracking to Improve Immersion in Reading," in *CERC*, 2017.

[11] T. Schlippe, M. Wölfel, and A. Stitz, "Generation of Text from an Audio Speech Signal," 2018, U.S. Patent 10043519B2.

[12] C. de Lacerda Pataca and P.D.P. Costa, "Speech modulated typography: towards an affective representation model," in *International Conference on Intelligent User Interfaces*, 2020, pp. 139–143.

[13] A. Bessemans, M. Renckens, K Bormans, E. Nuyts, and K. Larson, "Visual prosody supports reading aloud expressively," *Visible Language*, vol. 53, pp. 28–49, 12 2019.

References

[14] Q.V. Vy and D.I. Fels, "Using Placement and Name for Speaker Identification in Captioning," in *Computers Helping People with Special Needs*, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., 2010.

[15] A. Brown, R. Jones, M. Crabb, J. Sandford, M. Brooks, M. Armstrong, and C. Jay, "Dynamic Subtitles: The User Experience," in *TVX*, 2015.

[16] W. Fox, "Integrated titles: An improved viewing experience," *Eyetracking and applied linguistics*, 2016.

[17] R. McClarty, "Towards a multidisciplinary approach in creative subtitling," *MonTI: Monografías de Traducción e Interpretación*, pp. 133–153, 2012.

[18] G. El-Taweel, "Conveying Emotions in Arabic SDH: The Case of Pride and Prejudice," 2016, Master Thesis, Hamad Bin Khalifa University.

[19] J. Ohene-Djan, J. Wright, and K. Combie-Smith, "Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People," in *CVHI*, 2007.

[20] R. Rashid, J. Aitken, and D. Fels, "Expressing Emotions Using Animated Text Captions," 2006, Web Design for Dyslexics: Accessibility of Arabic Content.

[21] M. Gernsbacher, "Video Captions Benefit Everyone," *Policy Insights from the Behavioral and Brain Sciences*, vol. 2, pp. 195–202, 10 2015.

[22] A. Martone, C. Taskiran, and E. Delp, "Automated closed-captioning using text alignment," in *SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia*, 2004.

[23] G. Boulianne, J.-F. Beaumont, M. Boisvert, J. Brousseau, P. Cardinal, C. Chapdelaine, M. Comeau, P. Ouellet, and F. Osterrath, "Computer-assisted closed-captioning of live TV broadcasts in French," in *Interspeech*, 2006.

[24] K. Levin, I. Ponomareva, A. Bulusheva, G. Chernykh, I. Medennikov, N. Merkin, and N. Tomashenko, "Automated closed captioning for Russian live broadcasting," in *Interspeech*, 2014.

[25] S.-E. Tremblay A. Koul, R.G. Kulkarni, "Automated closed captioning using temporal data," 2015, U.S. Patent 9922095B2.

[26] J. Neves, "The (Very) Basics of Enriched Subtitles (for Deaf and Hard of Hearing Audiences) V2," 2019, Hamad bin Khalifa University.

References

[27] J. Neves, "Audiovisual Translation: Subtitling for the Deaf and Hard of Hearing," 2005, PhD Thesis, Roehampton University.

[28] R. Bringhurst, "The Elements of Typographic Style," *Hartley and Marks Publishers*, vol. 3.2, pp. 55–56, 2008.

[29] G. Unger, "Wie man's liest," *Niggli Verlag*, pp. 63–65, 2006.

[30] S.K. Alessai, "WaveFont for Arabic Video Captioning," 2020, Master Thesis, Hamad bin Khalifa University.

[31] T. Schlippe, T. Nguyen, and S. Vogel, "Diacritization as a Translation Problem and as a Sequence Labeling Problem," in *AMTA*, 2008.