AI-based Visualization of Voice Characteristics in Lecture Videos' Captions

Tim Schlippe[1], Katrin Fritsche[2], Ying Sun[2], Matthias Wölfel[3]

[1] IU International University of Applied Sciences, Germany
[2] University Jena, Germany
[3] Karlsruhe University of Applied Sciences, Germany
tim.schlippe@iu.org

Abstract. More and more educational institutions are making lecture videos available online. Since 100+ empirical studies document that captioning a video improves comprehension of, attention to, and memory for the video [1], it makes sense to provide those lecture videos with captions. However, studies also show that the words themselves contribute only 7% and how we say those words with our tone, intonation, and verbal pace contributes 38% to making messages clear in human communication [2]. Consequently, in this paper, we address the question of whether an AI-based visualization of voice characteristics in captions helps students further improve the watching and learning experience in lecture videos. For the AI-based visualization of the speaker's voice characteristics in the captions we use the *WaveFont* technology [3,4,5], which processes the voice signal and intuitively displays loudness, speed and pauses in the subtitle font. In our survey of 48 students, it could be shown that in all surveyed categories—*visualization of voice characteristics*, *understanding the content*, *following the content*, *linguistic understanding*, and *identifying important words*—always a significant majority of the participants prefers the *WaveFont* captions to watch lecture videos.

Keywords: closed captions; subtitles; speech processing; natural language processing; digital humanities, AI in education.

1 Introduction

Access to education is one of people's most important assets, and ensuring inclusive and equitable quality education is goal 4 of United Nations' Sustainable Development Goals [6]. Distance learning, in particular, can create education in areas where there are no educational institutions or in times of a pandemic. There are more and more distance learning opportunities worldwide and challenges like the physical absence of the teacher and the classmates or the lack of motivation of the students are addressed with technical solutions like video conferencing systems [7] and gamification of learning [8]. The research area "AI in Education" addresses the application and

evaluation of artificial intelligence (AI) methods in the context of education and training [9]. For instance, it deals with sentiment analysis to classify students' comments [10], natural language processing based tutoring systems [11], automatic short answer grading [12,13], recommender systems [14] or conversational AI systems, which optimally prepare students for their exams [15,16]. To overcome the linear structure of presentations and video lectures, [17] proposes to automatically generate a dialog system from slide-based presentations which can dynamically adapt to student requests. The capacity of the system is still limited, but its usefulness has already been confirmed by learners and lecturers alike.

Even though AI-based approaches show considerable potential, video lectures—along with classroom and online presentations—remain one of the dominant methods for conveying information. Since many educational institutions—public and private—already provide lecture videos online, even small advances in improving the effectiveness of video lectures will have a big impact on learners' knowledge acquisition. As understanding the spoken word is not always optimal due to a noisy recording (lecturers are not professional sound engineers), noisy environment of the viewer (e.g., on a train), or due to temporary or permanent disabilities, subtitling and captioning are used[1]. In addition, more than 100 empirical studies document that captioning a video improves comprehension of, attention to, and memory for the video [1]. However, conventional captions and subtitles have not evolved for decades. They still reflect *what* is spoken—not *how* it is spoken—i.e., no information about loudness, intonation, pauses, lengths, and emotions. We believe that potential is not being exploited here, as studies show that the words themselves contribute only 7% and how we say those words with our tone, intonation, and verbal pace contributes 38% to making messages clear in human communication [2]. Consequently, in this paper, we address the question of whether an AI-based visualization of voice characteristics in captions helps students further improve the watching and learning experience in lecture videos.

In the next section, we will present the latest approaches of other researchers to captioning and subtitling, as well as to the visualization of non-textual information, such as an utterance's emotion or prosody. In Section 3 we will introduce our technology *WaveFont* which we use to map acoustic features to font characteristics. Section 4 will describe the experimental setup for our user study. The study and the results are outlined in Section 5. We will conclude our work and suggest further steps in Section 6.

[1] In this research we refer to interlingual translation as *subtitles* and transcription in the same language as *captions*.

2 Related Work

Several studies show that educational videos should meet some requirements—be of shorter length, contain high quality image and text components, etc. [18,19]. The effect of captions and subtitles in those videos is described as supporting the understanding of thematic content as well as improving literacy and language skills [20,21,22,23]. [24] and [25] report that particularly in the university context students prefer videos with captions.

Studies which focus on captioning and subtitling in general are concerned with their placement and design [26,27]. An eye-tracking study indicates that these parameters can affect reading time and the visual perception of the image [28]. Traditional captions and subtitles are limited to telling the audience *what* is merely being said instead of *how* it is being said [29]. These methods do not present information beyond verbatim dialogue such as emotional expressions [30] and can lead to communication problems for the receiving audience [31]. Consequently, some studies assert the benefits of captions for the viewers, to make the material more understandable to them [32]. [3] state that creative captions and subtitles can benefit a wide range of people, not only deaf and hard-of-hearing. [33] investigates the use of creative subtitles through emojis and emoticons as well as reports that its use furthers the function of standard subtitling, allowing for tone of voice and emotions to be conveyed to the target audience.

To improve comprehension and to include non-textual information, such as emotion or prosody in an utterance, into a visual representation, [3] propose Voice-Driven Type Design (VDTD). It adjusts the shape of each single character according to particular acoustic features in the spoken reference. The motivation of a phoneme-to-grapheme adaptation is to better represent the characteristics of *how* it has been spoken besides *what* has been spoken. VDTD maps the three acoustic properties loudness, speed and pitch to the vertical stroke weight, horizontal stroke weight, and character width. [31] investigate how visual coding of prosody (bolded if louder, squished—what we refer to as narrow—if faster, etc.) can help children improve reading prosody. They found that coding verbal information can create an intuitive representation of speech's expressiveness. [34] exploit variable font technology to visualize voice characteristics at the syllable level and use letter slant to indicate prosody. They report that "participants' responses are highly consistent, indicating that it is indeed plausible to use typographic modulations as a way of representing speech expressiveness, or simply prosody". In another study [35], they report that when an example of their voice-modulated typography was shown along with two alternative sounds, participants correctly identified the original sound with an accuracy of 65%.

To the best of our knowledge, the first technology to visualize voice characteristics in captions is *WaveFont* [4,5]. *WaveFont* uses methods from automatic speech recognition, signal processing, machine learning, subtitling and typography to render

characteristics automatically and intuitively from the voice in captions. In the next section we will motivate our visualization and present our pipeline to generate *WaveFont* captions.

3 AI-based Visualization of Voice Characteristics in Captions

Our *WaveFont* visualization of the voice characteristics for video captions is at the word level, i.e., for each word, the average values of volume and length are used to decide which font to use to represent the whole word. It was very important for us to present the characteristics of the voice as intuitively as possible so that the viewer does not have to think long about how to interpret the visualization. Our previous analyses have shown that with captions, the word level is preferred to the character or syllable level since viewers see each caption only briefly and therefore have only a short time to interpret the visualization. Furthermore, although a representation of pitch is possible with our technology, we omit it as an additional visualized voice feature in our captions. The reasons are that previous studies have demonstrated that there is no clear agreement about the visualization of pitch in the typeface, pitch plays a subordinate role to the viewers compared to loudness and speed, and we do not want to overwhelm them with information. Consequently, we investigated different mappings and decided to map the voice to the character shape as follows:

- *Loudness*: Producing loudness in speech amplifies the signal and is usually used to attain the attention of a listener. To have the attention of the reader, bolder text is commonly used since it makes it easier and more efficient to scan the text and recognize important keywords [36]. Therefore, we use a thin font for quieter words and a bold font for louder words.
- *Speed*: The processes of information transfer with speech and reading happens within a time period. A reader usually jumps from a part of a word to the next part of a word [37]. Increasing the character width extends this scanning process of the eyes. Thus, we map the speed of the utterance to the character width: We use a narrow font for fast words and a wide font for slow words.

Our mapping is universally understood across cultures, while this is not the case for emojis and emoticons which may convey several meanings [33]. For aesthetic reasons the different fonts are chosen from the same font family. On the one hand, we aim not to have too extreme differences between the fonts, so that the typeface does not look too restless. On the other hand, the fonts need to be different enough to be easily recognized—even by inexperienced viewers on a small screen.

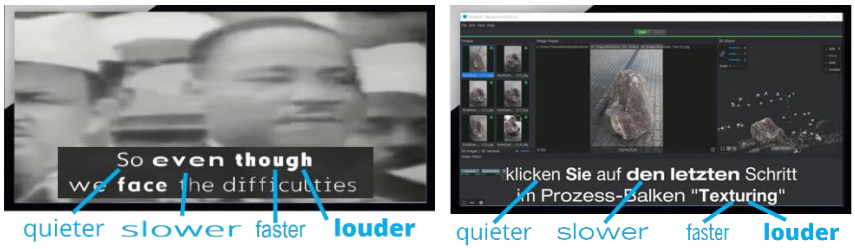


Fig. 1. *WaveFont* captions.

Figure 1 shows excerpts of Martin Luther King Jr.'s speech and of a German lecture video with *WaveFont* captions. The combination of the two visualizations for loudness and speed results in four classes. Figure 2 summarizes the mapping of the acoustic characteristics *loudness* and *speed* to its visual representations *stroke weight* and *character width* in our four classes.

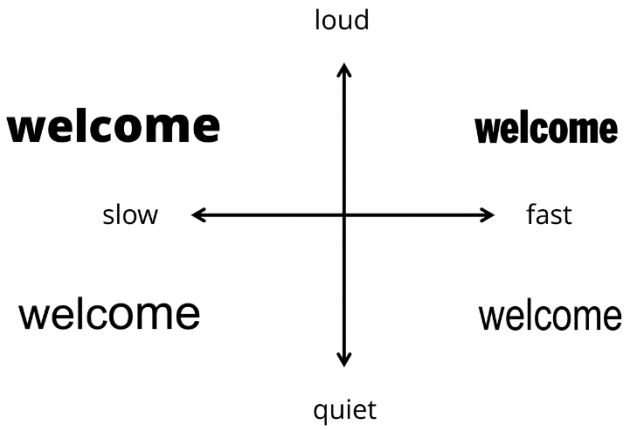


Fig. 2. Mapping speech characteristics on text formatting.

To generate *WaveFont* captions, we use a video and the corresponding caption file as input and apply the following steps [5]:

1. *Extraction of the audio track* from the video.
2. *Segmentation* of the audio track *into smaller audio files* containing spoken utterances based on the time information (start time and end time) of each caption.
3. For each audio speech segment: *Automatic forced alignment* process, which takes the text transcription of the audio speech segment and provides the start and end time of each word in the speech segment.
4. *Acoustic feature extraction*: Provides feature values of loudness and speed of each word. Loudness is based on the signal power. The feature values which

represent the speed of each spoken word are computed based on the number of characters and the duration of the uttered word.

5. *Mapping of acoustic features to font classes* based on thresholds.
6. *Type design*: Based on the content of the original subtitle file a new subtitle file is produced that contains font definitions according to the mapped font classes.

A detailed technical description is given in [3,4]. A benefit is that our technology can be ported to new languages and writing systems. For example, in [5], we describe how we adapted the system to Arabic.

4 Experimental Setup

In this section we describe the experimental setup of our study, which we conducted with a questionnaire.

4.1 Study Design

In order to ask questions about the two types of captions, we showed the participants footage from lectures with standard captions and with *WaveFont* captions at the beginning of the questionnaire. To get a representative lecture video for this purpose, we created a 1:46 minute compilation of excerpts from 7 lecture videos which were provided by the *Digital4Humanites* project[2]. The selected video snippets meet different criteria in terms of format and content: They are screencasts, slidecasts or animation videos and either address a theoretical treatise, the use of a software or an application. Six of the videos are in German and one in English. Furthermore, they cover various (digital) humanities subject areas: digital image measurement, museum data research, linguistics, 3D digital reconstruction, and general data-based skills for students. Our video compilation consists of an upper part where the lectures' video snippets are shown with standard captions and a lower part where the same snippets are shown simultaneously with *WaveFont* captions. This has the following advantages: (1) In contrast to having two separate videos, we exclude that participants watch the video with standard captions longer than the video with *WaveFont* captions or vice versa, which could influence the feedback. (2) The participants do not have to watch our video compilation twice, which would prolong the participation in our survey and could possibly lead to a decrease in the participants' motivation. (3) We enable a direct comparison of both visualizations.

In order to guarantee the participants the sole focus on the captions' visualization and a fair comparison, it was important to us that the standard captions and the *WaveFont* captions contain the same captions in terms of content and that these were created according to the best possible subtitle standards. Consequently, when we created the captions, we sticked to the German subtitle standards from ARD and ZDF[3]

[2] BMBF funding number: 16DHB3006; running time 1.1.2020–31.12.2022.
[3] http://www.untertitelrichtlinien.de

for the captions of the German lecture videos and to the BBC Subtitle Guidelines[4] for the English lecture video.

Our questionnaire was provided in German and English and consisted of a section where we asked for socio-demographic data, a section with general questions about the use of captions, and a section where we asked participants to compare standard and *WaveFont* captions in the categories *visualization of voice characteristics*, *understanding the content*, *following the content*, *linguistic understanding*, and *identifying important words*. The participants evaluated the questions with a score. The score range follows the rules of a forced choice Likert scale, which ranges from (1) *strongly disagree* to (5) *strongly agree*.

4.2 Participants

48 participants (23 female, 23 male, 2 diverse) filled out our questionnaire. The participants were students or former students at public or private universities and technical colleges, most of whom were between 18 and 44 years old and all participated free of charge. They represent students from a variety of disciplines, such as law, economics and social sciences, engineering, humanities, design and art as well as mathematics and natural sciences. Most of them have a high proficiency in the German language. But some participants are foreign students who do not speak German that well but are enrolled in German universities and listen to German lectures. 50% accessed our questionnaire via laptop or PC, 50% with their smartphones. We appreciate these distributions as it was important to us to get feedback from different people representing the target group who watch lecture videos.

5 Experiments and Results

In this section, we will describe the results of our study in which we compared standard captions to *WaveFont* captions with regard to the categories *visualization of voice characteristics*, *understanding the content*, *following the content*, *linguistic understanding*, and *identifying important words*. In addition, we will investigate for which applications standard and *WaveFont* captions have potential.

5.1 Visualization of Voice Characteristics

After the participants watched our 1:46 minute compilation of lecture videos, we asked them in our questionnaire if standard captions (*standard*) or *WaveFont* captions (*WaveFont*) visualize the characteristics of the lecturer's voice (e.g., loudness, lengths, pauses) better for them. Figure 3 illustrates their feedback. The blue pieces in the pie chart represent the proportion of participants who find *standard* better for this category. The green pieces represent the proportion who prefer *WaveFont*. As demonstrated in the figure, exactly half of the participants state that *WaveFont* shows them the

[4] https://bbc.github.io/subtitle-guidelines

characteristics of the voice *much better*. In addition, 27% indicate that *WaveFont* demonstrates these properties *better*. 15% let us know that *standard* visualizes the voice characteristics *better* and 0% that it shows these characteristics *much better*. 8% did not indicate a favorite method. Totally, in this category, 63% are more convinced of *WaveFont* than of *standard*.

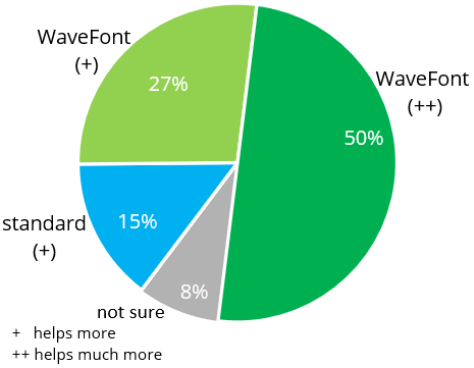


Fig. 3. Visualization of Voice Characteristics.

To get a better understanding if participants understand the concept of *WaveFont*, we asked if the participants find the visualization of *loudness*, the visualization of *speed* and finally the joint visualization of loudness and speed (*loudness+speed*) comprehensible. As illustrated in Figure 4, the visualization of *loudness* (3.90) was very well understood. Participants had more trouble understanding the visualization of *speed* (3.54), which is above the average score, indicating that it is still an adequate representation. The joint representation (*loudness+speed*) reduced comprehension (3.38).

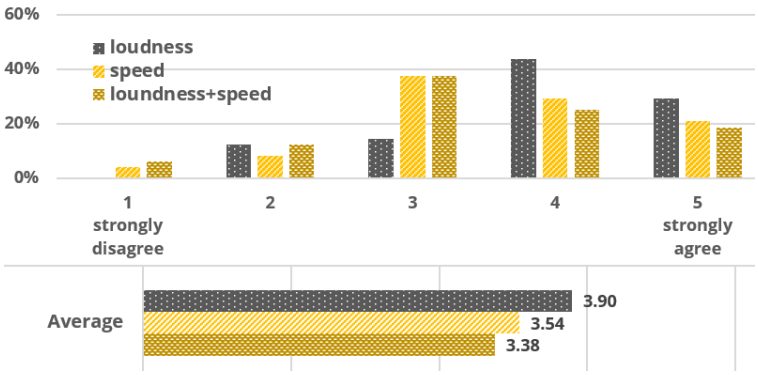


Fig. 4. Comprehensibility of *loudness*, *speed*, and *loudness+speed*.

However, when we asked if the participants agree that they assume to understand *WaveFont* with more practice even better, the majority agrees with an average of 4.19 in our 5-Likert scale as shown in Figure 5.

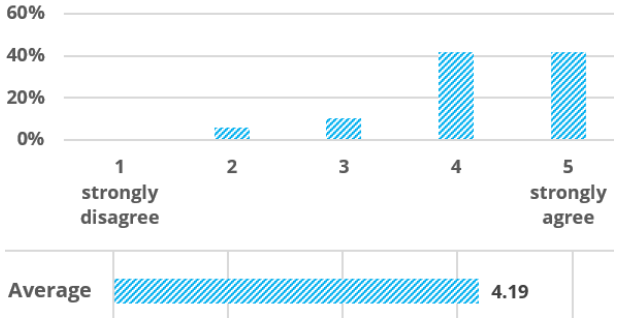


Fig. 5. Assumption of improved *WaveFont* comprehension with more practice.

In free text fields of the questionnaire, 4 students suggested using emojis and colors for the visualization of voice characteristics as well as including the lecturer's talking head (picture-in-picture). We do not consider these suggestions useful for the following reasons: While the representation of loudness with the stroke weight and of speed with the character width is intuitive [36,37], the interpretation of emojis depends on the cultural backgrounds, context, and individual characteristics [28]. The use of different colors in captions and subtitles already indicates different speakers in the BBC Subtitle Guidelines for English. In addition, the lecturer's talking head cannot always be faded in, e.g., not if it obscures important visual components or if the screen is too small.

5.2 Preferences

Figure 6 shows the participants' preferences with regard to captioning in lecture videos. For most students, *WaveFont* is the first choice. *standard* is chosen as second priority and no captions as third priority. The fact that more students prefer captions than no captions confirms the findings on educational videos described in [1]. The fact that the students favor *WaveFont*, which is more differentiated compared to *standard*, indicates that they like the advantages of visualizing loudness and speed using the typeface. The results of our more detailed analysis of acceptance are described in Section 5.3–5.6.

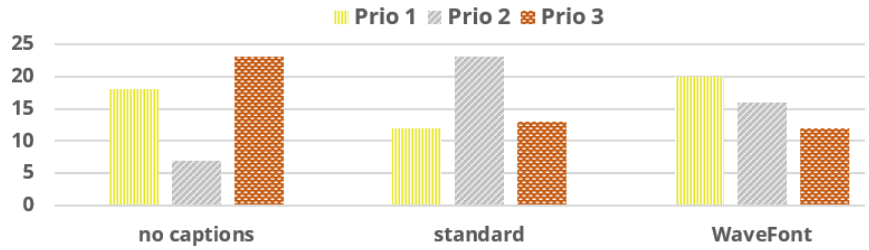


Fig. 6. Preferences in terms of captions in lecture videos (1st, 2nd, 3rd priority).

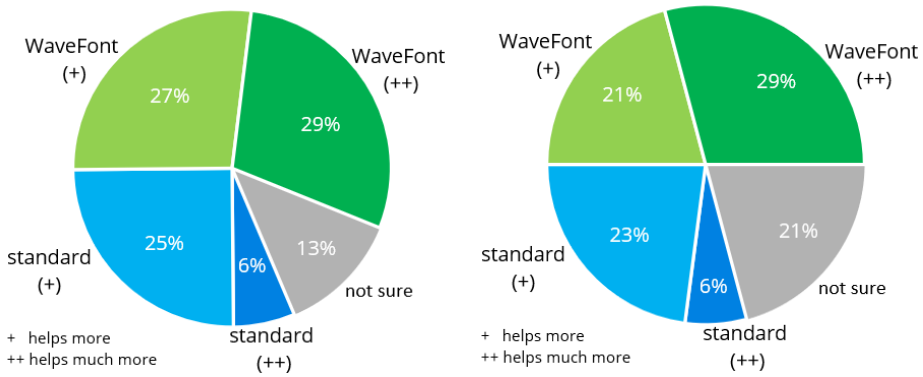
In addition to the general preference, we asked under which conditions the students find the use of *standard* or *WaveFont* in lecture videos particularly important. We received feedback that *standard* and *WaveFont* are particularly preferred when the sound quality of the video is poor, when there is background noise, when the lecturer speaks a language of which the students are not native speakers, and when the lecturer has a poor speaking style.

5.3 Understanding the Content

Since in lectures it is important that students understand the content as best as possible, we asked the participants if *standard* or *WaveFont* helps them *more* or even *much more* to understand the general content of the lecture video and specifically what the lecturer is saying in terms of content.

As shown in Figure 7 (a), 25% reported that *standard* helps them *more* with the general content understanding and 6% indicated that *standard* helps them even *much more*. In contrast, 27% ticked off that *WaveFont* helps them with understanding the content *more* than *standard* and even 29% that *WaveFont* helps them *much more*. 13% were not sure. In total, *WaveFont* convinced 56% of the participants in this category which is 25% more than those who think that *standard* is more helpful in this category.

As visualized in Figure 7 (b), more participants (21%) were not sure about the second question. But again, the majority (21%+29%) voted that *WaveFont* helps them *more* or *much more* than *standard* to better understand what the lecturer is saying in terms of content which is 21% more than those who voted for *standard*.



(a) understand lecture videos' content (b) what lecturer is saying about content

Fig. 7. Content understanding.

5.4 Linguistic Understanding

In addition to content understanding, it was important for us to find out which method helps the students more in terms of linguistic understanding. We wanted to figure out whether *standard* and *WaveFont* help students better grasp linguistic aspects of the lecturer's spoken language, such as grammar and pronunciation. Figure 8 demonstrates

that in this category even more welcome *WaveFont* with 63%. The support provided by the visualization of the voice characteristics is particularly evident in the fact that 38%—i.e., 9% more than with *content understanding*—report that *WaveFont* helps them *much more*. Only 27% (23%+4%) report that *standard* helps them *more* and *much more*. 10% choose neither *standard* nor *WaveFont*.

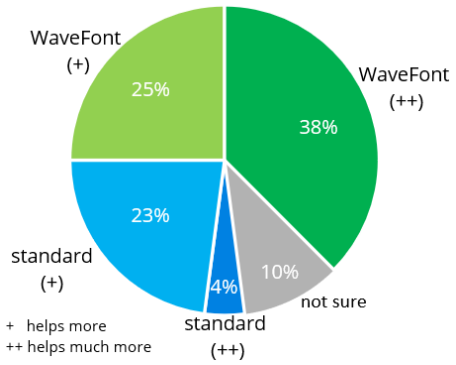


Fig. 8. Linguistic understanding.

5.5 Following the Content

Figure 9 illustrates our evaluation of whether *standard* or *WaveFont* helps them *more* or even *much more* to follow the content of the lecture. This time, slightly more participants reported that *standard* helps them follow the content *more* (27%) and *much more* (10%) than *WaveFont*. However, 27% indicated that *WaveFont* helps them with understanding the content *more* than *standard* and the same percentage that *WaveFont* helps them *much more*. 8% were not sure. Again, *WaveFont* convinced with 54% more than half of the participants in this category which is 17% more than those who prefer *standard* in this category.

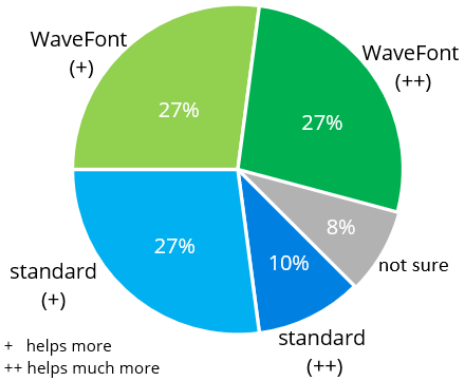


Fig. 9. Following the content.

5.6 Identifying Important Words

Furthermore, we investigated if our AI-driven *WaveFont* technology helps students better recognize words which are important to the lecturer. As illustrated in Figure 10, the support provided by the visualization of the speaker's voice characteristics is particularly evident in the large majority of 83% reporting that *WaveFont* helps them *more* (27%) and *much more* (56%). Only 10% report that *standard* helps them *more* and 0% that it helps *much more*. Only 6% were not sure. In this category *WaveFont* outperforms *standard* by 73% which is significantly better than in the other categories.

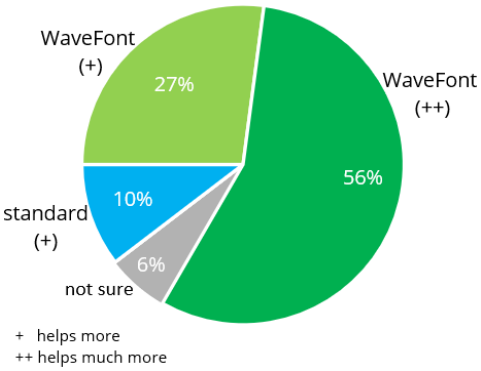


Fig. 10. Identifying important words.

5.7 Further Applications

Finally, we wanted to find out in which other applications the use of *WaveFont* has potential. When asked where the participants would like to see *WaveFont*, we got different answers as visualized in Figure 11. Use cases where more than 30% of the participants agree are: Live broadcasts, social media, video-on-demand, and videos on websites. In a past study in Arab countries, we also asked this question [5]. In that study, more than 30% of the participants agreed on the following use cases: Video-on-demand, TV, social media, live broadcasts, and TV sets in public places. This shows that there are similarities in the desired use cases between a group of people who are oriented towards German culture and a group of people who are more oriented towards Arabic culture, but also cultural differences.

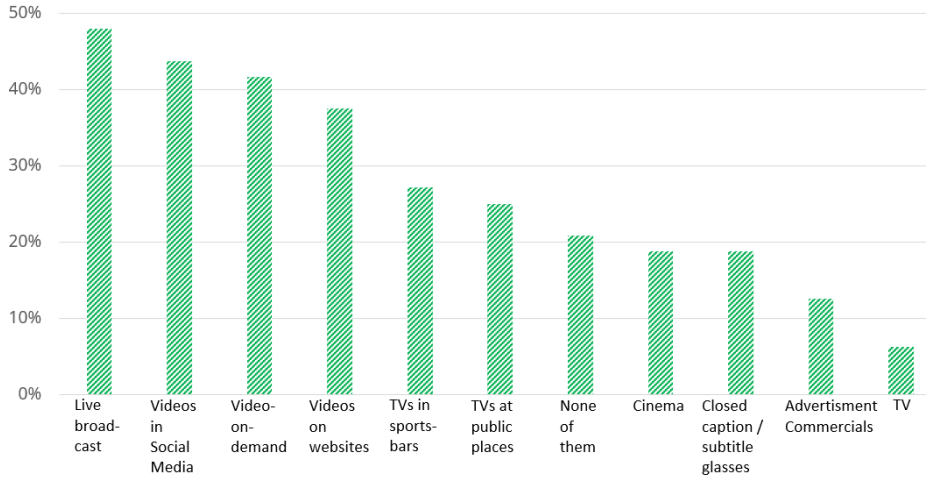


Fig. 11. Applications.

6 Conclusion and Future Work

In this paper, we have demonstrated that an AI-based visualization of voice characteristics in captions helps students improve the viewing and learning experience in lecture videos. For the AI-based visualization of the lecturer's voice characteristics in captions, we used our *WaveFont* technology [3,4,5], which 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 since it helps students visualize speech features, understand the content, follow the content, in linguistic understanding, and identify important words. When we asked if the participants agree that they assume to understand *WaveFont* with more practice even better, the majority agrees with an average of 4.19 in our 5-Likert scale.

Based on this good prediction, we would like to analyze the learning effect in future work in more detail, e.g., with time measurements, measuring cognitive load, eye tracking and with targeted questions about the content of video lectures. Thereby, we also plan to investigate the effect of language acquisition for non-native speakers. Moreover, it is interesting to analyze the effect of *WaveFont* captions on learning styles. On the one hand, visual learners are good at using vision to obtain information, and their information processing channels tend to use visual channels and are more inclined to use pictures to represent information and thoughts. Verbal learners, on the other hand, are good at using auditory information to obtain information, and their information processing channels are more likely to use auditory channels and more inclined to use words to present information and thoughts [39].

References

1. Gernsbacher M.A.: Video Captions Benefit Everyone. Policy Insights from the Behavioral and Brain Sciences, 2(1), 195–202 (2015)
2. Marteney, J.: Verbal and Nonverbal Communication. ASCCC Open Educational Resources Initiative (OERI). https://socialsci.libretexts.org/@go/page/67152 (2020)
3. Wölfel, M., Schlippe, T., Stitz, A.: Voice Driven Type Design. International Conference on Speech Technology and Human-Computer Dialog (SpeD), Bucharest, Romania (2015)
4. Schlippe, T., Wölfel, M., Stitz, A.: Generation of Text from an Audio Speech Signal, U.S. Patent 10043519B2 (2018)
5. Schlippe, T., Alessai, S., El-Taweel, G., Wölfel, M., Zaghouani, W.: Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic, International Conference on Cyberworlds (CW 2020), Caen, France (2020)
6. United Nations: Sustainable Development Goals: 17 Goals to Transform our World, https://www.un.org/sustainabledevelopment/sustainable-development-goals (2021)
7. Correia, A.P., Liu, C., Xu, F.: Evaluating Videoconferencing Systems for the Quality of the Educational Experience. Distance Education 41, 4, 429–452 (2020)
8. Koravuna, S., Surepally, U.K.: Educational Gamification and Artificial Intelligence for Promoting Digital Literacy. Association for Computing Machinery, New York, NY, USA (2020)
9. Chen, L., Chen, P., Lin, Z.: Artificial Intelligence in Education: A Review. IEEE Access 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510 (2020)
10. Rakhmanov, O., Schlippe, T.: Sentiment Analysis for Hausa: Classifying Students' Comments. The 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL 2022). Marseille, France (2022)
11. Libbrecht, P., Declerck, T., Schlippe, T., Mandl, T., Schiffner, D.: NLP for Student and Teacher: Concept for an AI based Information Literacy Tutoring System. The 29th ACM International Conference on Information and Knowledge Management (CIKM2020). Galway, Ireland (2020)
12. Sawatzki, J., Schlippe, T., Benner-Wickner, M.: Deep Learning Techniques for Automatic Short Answer Grading: Predicting Scores for English and German Answers. The 2nd International Conference on Artificial Intelligence in Education Technology (AIET 2021), Wuhan, China (2021)
13. Schlippe, T., Sawatzki, J.: Cross-Lingual Automatic Short Answer Grading. The 2nd International Conference on Artificial Intelligence in Education Technology (AIET 2021), Wuhan, China (2021)
14. Bothmer, K., Schlippe, T.: Investigating Natural Language Processing Techniques for a Recommendation System to Support Employers, Job Seekers and Educational Institutions. The 23rd International Conference on Artificial Intelligence in Education (AIED) (2022)
15. Bothmer, K., Schlippe, T.: Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-based Recommendation System. In Proceedings of The Learning Ideas Conference 2022 (15th annual conference), New York, New York, 15-17 June (2022)
16. Schlippe, T., Sawatzki, J.: AI-based Multilingual Interactive Exam Preparation. In The Learning Ideas Conference 2021 (14th annual conference). ALICE - Special Conference Track on Adaptive Learning via Interactive, Collaborative and Emotional Approaches. New York, USA. https://doi.org/10.1007/978-3-030-90677-1_38 (2021)
17. Wölfel, M.: Towards the Automatic Generation of Pedagogical Conversational Agents from Lecture Slides, International Conference on Multimedia Technology and Enhanced Learning (2021)

18. Ou, C., Joyner, D. A., Goel, A. K.: Designing and Developing Video Lessons for Online Learning: A Seven-Principle Model. Online Learning, 23(2), 82–104 (2019)
19. Wang, J., Antonenko, P., Dawson, K.: Does Visual Attention to the Instructor in Online Video Affect Learning and Learner Perceptions? An Eye-Tracking Analysis. Computers & Education, 146. https://doi.org/10.1016/j.compedu.2019.103779 (2020)
20. Perego, E., Del Missier, F., Porta, M., Mosconi, M.: The Cognitive Effectiveness of Subtitle Processing. Media Psychology, 13, 243–272. (2010)
21. Linebarger, D. L.: Learning to Read from Television: The Effects of Using Captions and Narration. Journal of Educational Psychology, 93, 288–298 (2001)
22. Bowe, F. G., Kaufman, A.: Captioned media: Teacher Perceptions of Potential Value for Students with No Hearing Impairments: A National Survey of Special Educators. Spartanburg, SC: Described and Captioned Media Program (2001)
23. Guo, P. J., Kim, J., Rubin, R.: How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. L@S '14: Proceedings of the first ACM conference on Learning. March 2014, 41-50. https://doi.org/10.1145/2556325.2566239 (2014)
24. Alfayez, Z. H.: Designing Educational Videos for University Websites Based on Students' Preferences. Online Learning, 25(2), 280-298 (2021)
25. Persson, J. R., Wattengård, E., Lilledahl, M. B.: The Effect of Captions and Written Text on Viewing Behavior in Educational Videos. International Journal on Math, Science and Technology Education, 7(1), 124–147 (2019)
26. Vy, Q.V., Fels, D.I.: 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)
27. Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., Jay, C.: Dynamic Subtitles: The User Experience. TVX (2015)
28. Fox, W.: Integrated titles: An Improved Viewing Experience. Eyetracking and Applied Linguistics (2016)
29. Ohene-Djan, J., Wright, J., Combie-Smith, K.: Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People. CVHI (2007)
30. Rashid, R., Aitken, J., Fels, D.: Expressing Emotions Using Animated Text Captions. Web Design for Dyslexics: Accessibility of Arabic Content (2006)
31. Bessemans, A., Renckens, M., Bormans, K., Nuyts, E., Larson, K.: Visual Prosody Supports Reading Aloud Expressively. Visible Language, vol. 53, pp. 28–49 (2019)
32. Gernsbacher, M.: Video Captions Benefit Everyone. Policy Insights from the Behavioral and Brain Sciences, vol. 2, pp. 195–202 (2015)
33. El-Taweel, G.: Conveying Emotions in Arabic SDH: The Case of Pride and Prejudice. Master Thesis, Hamad Bin Khalifa University (2016)
34. de Lacerda Pataca, C., Costa, P.D.P.: Speech Modulated Typography: Towards an Affective Representation Model. In International Conference on Intelligent User Interfaces, pp. 139–143 (2020)
35. de Lacerda Pataca, C., Dornhofer Paro Costa, P.: Hidden Bawls, Whispers, and Yelps: Can Text be Made to Sound More than just its Words?. arXiv:2202.10631 (2022)
36. Bringhurst, R.: The Elements of Typographic Style. Hartley and Marks Publishers, vol. 3.2, pp. 55–56 (2008)
37. Unger, G.: Wie man's liest. Niggli Verlag, pp. 63–65 (2006)
38. Bai, Q., Dan, Q., Mu, Z., Yang, M.: A Systematic Review of Emoji: Current Research and Future Perspectives. Front. Psychol. 10:2221. doi: 10.3389/fpsyg.2019.02221 (2019)
39. R: Rayner, S. G.: Cognitive Styles and Learning Styles. In, J. D. Wright, (Ed.). International Encyclopedia of Social and Behavioral Sciences (2nd edition), Vol 4, pp. 110–117. Oxford: Elsevier (2015)