Laugh at Your Own Pace: Basic Performance Evaluation of Language Learning Assistance by Adjustment of Video Playback Speeds Based on Laughter Detection

Naoto Nishida
University of Tsukuba
Ibaraki, Japan
nawta1998@gmail.com

Hinako Nozaki
University of Tsukuba
Ibaraki, Japan
nozaki@iplab.cs.tsukuba.ac.jp

Buntarou Shizuki
University of Tsukuba
Ibaraki, Japan
shizuki@cs.tsukuba.ac.jp

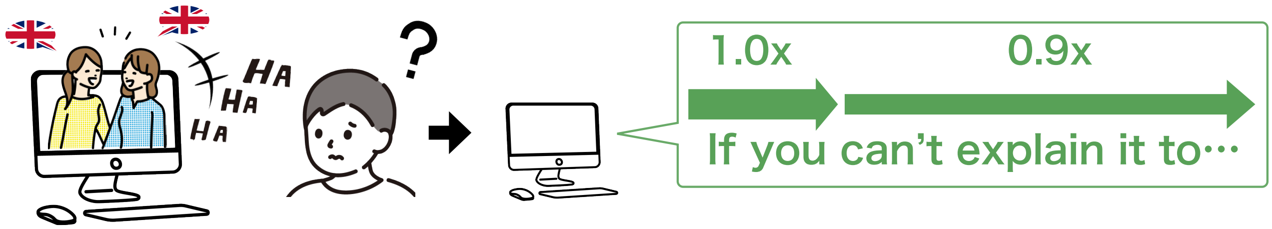

Figure 1: System overview. If learners do not laugh at the punchlines of videos, our system judges that they do not comprehend the video content, and gradually decreases the playback speed. Conversely, the speed increases if learners laugh.

ABSTRACT

Among various methods to learn a second language (L2), such as listening and shadowing, *Extensive Viewing* involves learning L2 by watching many videos. However, it is difficult for many L2 learners to smoothly and effortlessly comprehend video contents made for native speakers at the original speed. Therefore, we developed a language learning assistance system that automatically adjusts the playback speed according to the learner's comprehension. Our system judges that learners understand the contents if they laugh at the punchlines of comedy dramas, and vice versa. Experimental results show that this system supports learners with relatively low L2 ability (under 700 in TOEIC Score in the experimental condition) to understand video contents. Our system can widen learners' possible options of native speakers' videos as *Extensive Viewing* material.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → *Education*.

KEYWORDS

Human-Computer Interaction, Computer-Assisted Language Learning, Extensive Viewing, Language, Learning, Facial Expression

L@S '22, June 1–3, 2022, New York City, NY, USA

ACM ISBN 978-1-4503-9158-0/22/06…$15.00
https://doi.org/10.1145/3491140.3528299

ACM Reference Format:
Naoto Nishida, Hinako Nozaki, and Buntarou Shizuki. 2022. Laugh at Your Own Pace: Basic Performance Evaluation of Language Learning Assistance by Adjustment of Video Playback Speeds Based on Laughter Detection. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22), June 1–3, 2022, New York City, NY, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491140.3528299

1 INTRODUCTION

Among various methods to learn a second language (L2), Renandya and Jacobs defined *Extensive Viewing* as "listening to large amounts of motivating and engaging materials which are linguistically appropriate over a while where they listen with a reasonable speed for general understanding, with a focus on meaning rather than form" [39]. Ivone and Renandya further defined it as "to be exposed to a large amount of easily comprehensible and enjoyable materials presented in the target language over an extended period" as a kind of *Extensive Listening* [24]. *Extensive Viewing* has various benefits as a learning method; learners can improve their listening fluency [7, 39], increase their listening vocabulary [15, 29, 30], and get used to liaisons, slang, and spoken forms of the target language [3, 40]. Besides, it is also superior to akin *Extensive Learning* methods because engaging and real-life videos can maintain learners' motivation towards learning [52], learners can start learning whenever they want [8, 47], and captions can aid the comprehension for the contents [21, 28, 36]. Therefore, *Extensive Viewing* recently gained the attention of language learners as a valuable and enjoyable method [24].

However, it is difficult for many L2 learners to view the learning materials at the rate at which native speakers do because of their undeveloped reading and listening knowledge of the target language [5, 47]. As such, they usually face scenes where they cannot fully understand the content at the original playback speed.

As easily comprehensible material is essential for *Extensive Viewing* [24], video materials for native speakers (e.g., drama series) at the original speed are not suitable for L2 learners.

On the other hand, making L2 learners manipulate interfaces to change the playback speed is also problematic. This is because continuous manipulation over a long period can make learners less motivated [17], which is not suitable for *Extensive Viewing* as a method that takes over months or years [32, 39, 47].

Therefore, we created a language learning assistance system that enables L2 learners to effortlessly engage in long-lasting *Extensive Viewing* of videos made by native speakers. This system can automatically adjust the playback speed of videos depending on the comprehension of each L2 learner based on their facial expressions (Figure 1).

For this purpose, we first examined if learners act differently depending on their comprehension. From our preliminary study, we found that if learners laugh at the punchlines of videos, then they comprehend the content, and vice versa. We then proved that our proposed system is adequate for L2 learners with low proficiency. Our research can broaden learners' possible choices of materials for *Extensive Viewing*.

Our contributions can be summarized as follows:

- We found appropriate indicators that suggest learners' comprehension at a specific time while viewing videos (i.e., gaze, facial expressions, and other gestural behaviors).
- We developed a novel method to support L2 learners by adjusting the playback speed according to the comprehension of each learner.
- We evaluated the system via an experiment, revealing that it is beneficial for learners whose L2 skills are relatively low (under 700 in TOEIC score in the experimental condition).

2 RELATED WORK

2.1 Existing Computer-Assisted Language Learning Methods

In the context of *Extensive Viewing*, existing computer-assisted language learning methods mainly focus on developing the vocabulary sizes of L2 learners. For instance, Hu et al. and Sakunkoo et al. developed a system that enables learners to watch videos with interactively translated captions, where learners hover a mouse on unknown L2 words [22, 43]. Fujii and Rekimoto developed a system that detects English learners' proficiency level to suggest a moderate number of translations [17].

However, there are few systems to support L2 learners using playback speed control in *Extensive Viewing*. Besides, existing systems require a series of movements via mouse cursors from users to change the speed of videos [11, 22, 48]. *Extensive Viewing* requires continuous efforts towards learning from learners. Therefore, the workload necessary for viewing should be reduced as much as possible.

Thus, we developed a system that supports L2 learners through video speed control and requires less workload.

2.2 Relationship between Playback Speed and Learner Comprehension

In terms of foreign language learning, Blau and Sugai et al. confirmed that a slower speech rate facilitates listening comprehension among L2 learners [5, 49]. On the other hand, in terms of Massive Open Online Courses (MOOCs) in mother tongues, Kao et al. confirmed that sped up lecture videos boost MOOC learners' comprehension, while shortening the viewing time and raising the probability of completing the course [25, 27].

From these results, we induced that if learners cannot understand the video contents well, slowing the video content would be helpful because they can spend additional time processing the oral information, whereas if learners can understand them well, speeding up the video content would benefit learners because they can spare unnecessary time to watch and keep their concentration.

2.3 Reduction of Learners' Workload

For learners maintaining a daily habit of *Extensive Viewing*, the workload necessary for the manipulation should be reduced as much as possible. Fujii et al. and Song et al. used unconscious gaze movements or postures to predict learners' L2 proficiency levels [17, 45]. Similarly, Arakawa and Yakura used learners' unconscious reactions towards oral alterations to keep their attention during video-based learning, while maintaining the mental workload at a constant level [1].

Hence, we used learners' unconscious behavior to alleviate the workload necessary for learning. Specifically, we focused on whether learners laugh when they face the punchline scenes of video material.

2.4 Apparatus for *Extensive Viewing*

One of the aims of *Extensive Viewing* is to increase the L2 learning time outside of the classroom [39, 47]. In addition, *Extensive Viewing* requires long-lasting continuation on a daily basis [47]. Therefore, to reduce the workload required for the setup for learning anywhere and anytime, we should minimize the apparatus for *Extensive Viewing*.

Therefore, we examined features that we can capture via a built-in webcam of an ordinary laptop computer (MacBook Pro).

3 IMPLEMENTATION

Through our preliminary study with 5 participants (all males, M=23.6, SD=1.82, all of them are Japanese and learn English as a foreign language), we discovered that people laugh at the punchlines of content only when they understand the English video content well. Figure 2 presents the total Action Unit 14 (AU14) values according to the frame number of the experimental videos[1]. Action Units are measures to capture human facial movements by the appearance on their face [9, 12, 13]. AU14 corresponds to a movement of buccinator, which can be an index of smile and laugh [18, 50]. In the analysis of our preliminary study, we confirmed a significant difference in AU14 between when learners keep up with the contents and when they cannot ($p<.001$ in Mann-Whitney U test, and effect size 0.22 in

[1] https://drive.google.com/file/d/1dUrHqrXmw1APNtyVhdz-2tVeLOofoMbN/view?usp=sharing

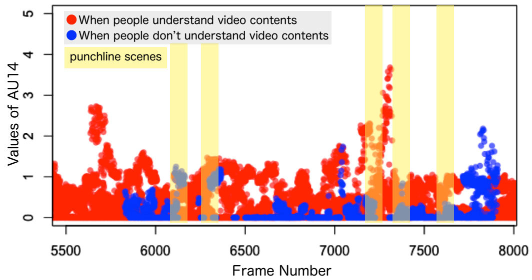

Figure 2: Values of Action Units 14 (*AU14*) of the 5 participants according to frame number. Red points represents the case when learners understand the contents. Blue points represents the case when they do not understand. Yellow ranges are punchline scenes in the experimental video. In the preliminary study, participants watched the video twice. First, they watched it normally; second, while watching they classified when they can keep up with the story and when they cannot.

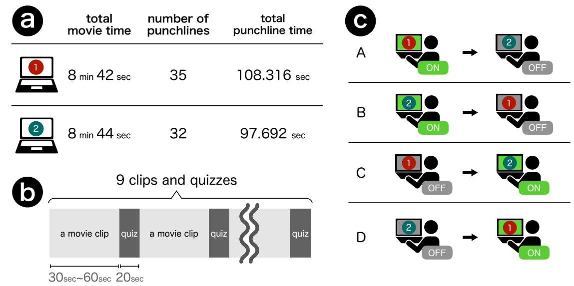

Figure 3: Control configuration: a) control in two videos, b) composition of a video, and c) control in groups.

Table 1: TOEIC scores in each group.

	TOEIC score					Mean (SD)
A	565	635	700	755	855	702.00 (111.2)
B	550	690	755	760	815	714.00 (101.8)
C	590	670	730	790	845	725.00 (99.87)
D	550†	670	680	785	880	713.00 (111.9)

† One person had not taken TOEIC. Thus, we hypothesized their score from their statement.

Hedges' g; we used Kolmogorov-Smirnov test for normality confirmation), whereas the other AUs, face tilt, and gaze direction did not show a significant difference. Some participants also said that if they cannot keep up with the story, their face gets strained, and they cannot laugh at the punchlines of the contents. Additionally, we prototyped a system that adjusts the playback speed of videos depending on the viewer's facial expressions and demonstrated it at a workshop in Japan [33]. We then ensured changing the playback speed could help learners comprehend the contents.

Based on the result of that preliminary study, we implemented a language learning assistance system that personalizes the playback speeds of video material based on laughter detection in the punchlines of the videos.

We used Python 3.7.4, Perception for Autonomous Systems (PAZ) library for facial expression detection, Selenium library for manipulation of the video player, and Laughter-Detection library for detecting the punchlines of video materials [2, 10, 19]. We defined a punchline as a scene where a laugh track is inserted, assuming that situation comedy (i.e., content with laugh tracks) is the de facto standard for *Extensive Viewing* [47, 52]. We chose video streaming sites (e.g., YouTube, Netflix) as a video player platform because these sites employ a time stretch algorithm, which provides a steady pitch regardless of playback speed changes [37].

This system decreases the playback speed by 0.1x when the user does not show a smile or laugh in punchline scenes in the videos, and increases the speed by 0.1x when the user shows a smile or laugh in punchline scenes. The playback speed ranges from 0.6x to 1.0x (original speed).

4 EXPERIMENT

We conducted an experiment to evaluate the system's basic performance to help learners understand videos while keeping the workload and enjoyment towards the content steady, which are essential factors for the continuation of *Extensive Learning*. We selected English as the target language because of its prevalence, making it easier to look for participants who learn English as a foreign language (EFL).

4.1 Participants

We recruited EFL learners whose English proficiency level is above B1 in CEFR score [34], which is 550 in TOEIC score [41]. This is because *Extensive Learning* is a method mainly for people who know at least 3000 basic words [38, 42, 51], and people whose English CEFR score is above B1 generally know more than 3000 words [4, 31, 46]. Twenty participants (4 females, 16 males, M=22.6, SD=1.59, all Japanese) participated in this experiment. We asked the participants to join this experiment with their ordinary eye conditions and make their faces easy to see, such as brushing up bangs, during the experiment. 8 participants joined the experiment with naked eyes, 5 participants with glasses, and 7 participants with contact lenses.

4.2 Apparatus and Location

We used a laptop PC (MacBook Pro 13 inch, Big Sur 11.6, 2.8 GHz quad-core Intel Core i7, 16 GB memory). The built-in webcam was a 720p FaceTime HD camera (1280 x 720 pixels). We avoided backlit conditions to capture the face clearly. The distance between a participant's face and the PC display was from 40 cm to 60 cm, which is the recommended distance of PC manufacturer [35].

4.3 Configuration

To reduce bias from the video itself, we made two similarly constructed videos (*V1*[2] and *V2*[3]). 578 words were spoken in *V1*, and 469 words were spoken in *V2*. Each video was composed of 9 clips

[2]https://drive.google.com/file/d/1LYx2rFDSLjlLNZo5Pa6qiUuhSweM_Y0W/view?usp=sharing
[3]https://drive.google.com/file/d/18cyV69l3x0ieA6xVUyg15MuHxcDOJKMz/view?usp=sharing

Table 2: Time spent on each process.

	answering time on *VT1*, *VT2*				reading quiz beforehand		viewing time with the system
	pre-viewing		post-viewing				
	VT1	*VT2*	*VT1*	*VT2*	*Q1*	*Q2*	
mean(s)	111.8	138.6	44.86	48.37	169.6	164.0	634.5
SD(s)	31.27	56.57	13.02	19.51	86.79	92.38	44.62

from a sitcom *Friends* and 9 corresponding comprehension quizzes about the clips (Figure 3 a and Figure 3 b). Each clip's length was approximately from 60 to 90 seconds, and each quiz's length was 20 seconds based on the example of listening quizzes of TOEIC and TOEFL [16, 44]. Because we compare the difference in comprehension between when a participant uses our system and when they do not, we divided the participants into 4 groups (Figure 3 c, Table 1). We allocated them according to their TOEIC score because the reliability of the test is high (0.90) [23, 26], meaning that its score almost certainly reflects learners' L2 ability. In addition, to reduce the effects of the participants' vocabulary ability, we aimed to ensure every word in the video was known. To this end, we let them read a series of quizzes shown in the videos and taught them the unknown vocabulary spoken in the videos. We also had them read a complex vocabulary list. We took the 20 most rarely seen words from each video with reference to the *Corpus of Contemporary American English* [14]. We ensured all participants knew the vocabulary in the videos and quizzes in our post-questionnaire.

4.4 Procedure

We take group A (Figure 3 c) as an example to explain the procedure. First, participants answered a vocabulary test (*VT1*) whose vocabulary appears in the following *V1*. The vocabulary test was conducted to teach participants the difficult words necessary to keep up with the story; therefore, we told participants the correct answer immediately after they finished answering. Second, we let them read a series of quizzes (*Q1*) to be presented in the quiz time in *V1* to ensure they had no vocabulary problems while answering. Third, participants viewed *V1* with our proposal system, answering *Q1* at the same time. Fourth, participants answered *VT1* again to be checked if they had no vocabulary issues while viewing *V1*. Fifth, participants answered NASA-TLX, an assessment tool that rates perceived workload [20]. After that, participants were given the same series of workflows again under the condition of *V2*, *VT2*, *Q2*, and without-system viewing instead of *V1*, *VT1*, *Q1*, and with-system viewing. Participants were given 5 minutes rest after answering NASA-TLX. At the end of the experiment, they answered a free answer sheet and SUS to assess usability [6]. We told the participants to write about any unknown words in the videos and quizzes. The whole experiment time was 64 minutes on average. Whether a participant used the system was single-blinded.

4.5 Results

The time spent on each process is shown in Table 2. The time participants viewed videos without the system was 523 seconds on average, meaning they spent 1.2 times longer than they viewed without the system.

Regarding the participants' comprehension, as shown in Figure 4, learners with low L2 proficiency (the lowest 2 people from each group) could answer both *V1* and *V2* more correctly with our system (*Q1*:p<.005 and *Q2*:p<.05, the effect sizes are 2.69 and 1.71

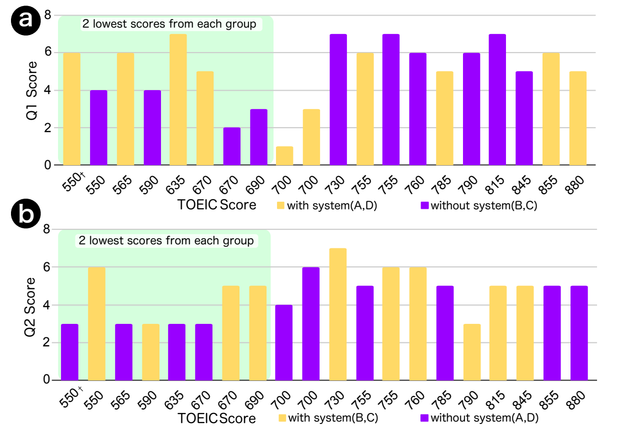

Figure 4: Relationship between TOEIC score and *Q1* (a) or *Q2* (b). Relatively, participants with low TOEIC score had a higher score when they used our system. The green area indicates the participants with the lowest and second lowest scores from each group.

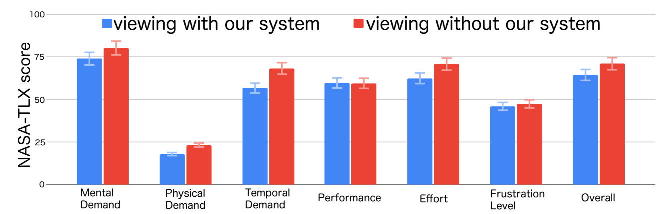

Figure 5: NASA-TLX score and its 10% error range. This shows that our system mostly does not worsen learners' mental workload. Besides, it significantly alleviates *Temporal Demand* and *Effort*.

respectively). This means that learners with low L2 proficiency levels can understand videos more with our system.

In terms of mental workload, NASA-TLX shows lower scores from perspectives other than *Performance* [20], showing significant differences in *Temporal Demand* and *Effort* (Figure 5).

As for system usability, SUS scored 73.75 on average (SD=14.66), which represents *Good* usability [6].

5 DISCUSSION & FUTURE WORK

From our experiment, we confirmed that our system enhanced L2 learners' comprehension of video content as a language learning material. The effect was particularly significant for learners with low proficiency (under 700 in TOEIC score in our experiment condition). We also proved that this system costs learners less mental workload and provides *Good* usability.

In future work, we will increase the number of participants with lower L2 proficiency (B1 or below in L2 proficiency). We also plan to discover more appropriate ways to extract learners' comprehensions. Besides, we will improve our system so that people can speed up the playback rate past 1.0x to skip any scene they can easily understand. This allows learners to spare time, leading to less workload for learning and the daily continuation of *Extensive Viewing*.

REFERENCES

[1] Riku Arakawa and Hiromu Yakura. 2021. Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 99, 15 pages. https://doi.org/10.1145/3411764.3445339

[2] Octavio Arriaga. 2021. Perception for Autonomous Systems. https://github.com/oarriaga/paz (Last Access Date:2022/4/6).

[3] Betül Bal-Gezegin. 2014. An Investigation of Using Video vs. Audio for Teaching Vocabulary. *Procedia - Social and Behavioral Sciences* 143, 3 (2014), 450–457. https://doi.org/10.1016/j.sbspro.2014.07.516

[4] Veronica Benigno and John De Jong. 2019. Linking vocabulary to the CEFR and the Global Scale of English: A psychometric model. In *Developments in Language Education: A Memorial Volume in Honour of Sauli Takala*, Ari Huhta, Gudrun Erickson, and Neus Figueras (Eds.). EALTA-European Association for Language Testing and Assessment, University of Jyvaskyla, Centre for Applied Language Studies, 8–29.

[5] Eileen Kay Blau. 1990. The Effect of Syntax, Speed, and Pauses on Listening Comprehension. *TESOL Quarterly* 24, 4 (1990), 746–753. https://doi.org/10.2307/3587129

[6] John Brooke. 1996. SUS: A "Quick and Dirty" Usability Scale. In *Usability Evaluation In Industry*. London: Taylor and Francis, 189–194.

[7] Anna CS Chang. 2016. Teaching L2 Listening: In and Outside the Classroom. In *English Language Teaching Today: Linking Theory and Practice*. Springer, 111–125.

[8] Hung chun Wang and Cheryl Wei yu Chen. 2020. Learning English From Youtubers: English L2 Learners' Self-regulated Language Learning on Youtube. *Innovation in Language Learning and Teaching* 14, 4 (2020), 333–346. https://doi.org/10.1080/17501229.2019.1607356

[9] Jeffrey Cohn, Zara Ambadar, and Paul Ekman. 2007. Observer-Based Measurement of Facial Expression with the Facial Action Coding System. In *The Handbook of Emotion Elicitation and Assessment*. Oxford University Press Series in Affective Science, 2–53.

[10] Software Freedom Conservancy. 2021. Selenium. https://www.selenium.dev/ja/ (Last Access Date:2022/4/6).

[11] Dioco. 2022. Language Reactor. https://www.languagereactor.com/.

[12] Paul Ekman. 1993. Facial Expression and Emotion. *American psychologist* 48, 4 (1993), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

[13] Paul Ekman and Wallace V Friesen. 1978. *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

[14] English-Corpora.org. 2021. Corpus of Contemporary American English. https://www.english-corpora.org/coca/(Last Access Date:2022/4/6).

[15] Yanxue Feng and Stuart Webb. 2020. Learning Vocabulary Through Reading, Listening, and Viewing: Which Mode of Input Is Most Effective? *Studies in Second Language Acquisition* 42, 3 (2020), 499–523. https://doi.org/10.1017/S0272263119000494

[16] The Institute for International Business Communication. 2022. TOEIC Listening & Reading Test. https://www.iibc-global.org/toeic/test/lr.html (Last Access Date:2022/4/6).

[17] Katsuya Fujii and Jun Rekimoto. 2019. SubMe: An Interactive Subtitle System with English Skill Estimation Using Eye Tracking. In *Proceedings of the 10th Augmented Human International Conference 2019*. Association for Computing Machinery, New York, NY, USA, Article 23, 9 pages. https://doi.org/10.1145/3311823.3311865

[18] Mehdi Ghayoumi and Arvind Bansal. 2015. Unifying Geometric Features and Facial Action Units for Improved Performance of Facial Expression Analysis. In *New Developments in Circuits, Systems, Signal Processing, Communications and Computers*. 259–266.

[19] Jon Gillick. 2021. Laughter-Detection. https://github.com/jrgillick/laughter-detection (Last Access Date:2022/4/6).

[20] Sandra Hart and Lowell Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter Hancock and Najmedin Meshkati (Eds.). Vol. 52. North–Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[21] Abdolmajid Hayati and Firooz Mohmedi. 2011. The Effect of Films With and Without Subtitles on Listening Comprehension of EFL Learners. *British Journal of Educational Technology* 42, 1 (2011), 181–192. https://doi.org/10.1111/j.1467-8535.2009.01004.x

[22] Sathaporn "Hubert" Hu and Wesley J. Willett. 2018. Kalgan: Video Player for Casual Language Learning. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188498

[23] Gwan-Hyeok Im and Liying Cheng. 2019. The Test of English for International Communication (TOEIC®). *Language Testing* 36, 2 (2019), 315–324. https://doi.org/10.1177/0265532219828252

[24] Francisca Ivone and Willy Renandya. 2019. Extensive Listening and Viewing in ELT. *Teflin Journal* 30, 2 (2019), 237–256. https://doi.org/10.15639/teflinjournal.v30i2/237-256

[25] Chen-Tai Kao, Yen-Ting Liu, and Alexander Hsu. 2014. Speeda: Adaptive Speed-up for Lecture Videos. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 97–98. https://doi.org/10.1145/2658779.2658794

[26] Frederic Kuder and Marion Richardson. 1937. The Theory of The Estimation of Test Reliability. *Psychometrika* 2, 3 (1937), 151–160. https://doi.org/10.1007/BF02288391

[27] David Lang, Guanling Chen, Kathy Mirzaei, and Andreas Paepcke. 2020. Is Faster Better? A Study of Video Playback Speed. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. Association for Computing Machinery, New York, NY, USA, 260–269. https://doi.org/10.1145/3375462.3375466

[28] Mina Lee, Beverly Roskos, and David R. Ewoldsen. 2013. The Impact of Subtitles on Comprehension of Narrative Film. *Media Psychology* 16, 4 (2013), 412–440. https://doi.org/10.1080/15213269.2013.826119

[29] Wang Li and Willy Renandya. 2012. Effective Approaches to Teaching Listening: Chinese EFL Teachers' Perspectives. *The Journal of Asia TEFL* 9, 4 (2012), 79–111. https://doi.org/10.1177/2158244020917393

[30] Ahmed Masrai. 2020. Can L2 Phonological Vocabulary Knowledge and Listening Comprehension be Developed Through Extensive Movie Viewing? The Case Of Arab EFL Learners. *International Journal of Listening* 34, 1 (2020), 54–69. https://doi.org/10.1080/10904018.2019.1582346

[31] James Milton and Thomaï Alexiou. 2009. Vocabulary Size and the Common European Framework of Reference for Languages. In *Vocabulary Studies in First and Second Language Acquisition*. Springer, 194–211. https://doi.org/10.1057/9780230242258

[32] Setsuko Mori. 2015. If You Build It, They Will Come: From a "Field of Dreams" to a More Realistic View of Extensive Reading in an EFL Context. *Reading in a foreign language* 27 (2015), 129–135.

[33] Naoto Nishida, Hinako Nozaki, and Buntarou Shizuki. 2021. Language Learning Assistance by Automatic Adjustment of Playback Speed Based on Facial Expression. In *29th Workshop on Interactive Systems and Software*. Japan Society for Software Science and Technology, Hamamatsu, Shizuoka, Japan. (in Japanese, non-archival article).

[34] Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://www.coe.int/en/web/common-european-framework-reference-languages.

[35] The College of Optometrists. 2022. Look After Your Eye. https://lookafteryoureyes.org/eye-care/screen-use/ (Last Access Date:2022/4/6).

[36] Elke Peters, Eva Heynen, and Eva Puimège. 2016. Learning Vocabulary Through Audiovisual Input: the Differential Effect of L1 Subtitles and Captions. *System* 63 (2016), 134–148. https://doi.org/10.1016/j.system.2016.10.002

[37] Pallavi Powale. 2017. Variable speed playback on mobile. https://www.english-corpora.org/coca/ (Last Access Date:2022/4/6).

[38] Geòrgia Pujadas and Carmen Muñoz. 2020. Examining Adolescent EFL Learners' TV Viewing Comprehension Through Captions And Subtitles. *Studies in Second Language Acquisition* 42, 3 (2020), 551–575. https://doi.org/10.1017/S0272263120000042

[39] Willy Renandya and George Jacobs. 2016. Extensive Reading and Listening in the L2 Classroom. In *English Language Teaching Today: Linking Theory and Practice*. Springer, Chapter 8, 97–110.

[40] Willy A Renandya and Thomas SC Farrell. 2011. 'Teacher, the tape is too fast!' Extensive Listening in ELT. *ELT journal* 65, 1 (2011), 52–59. https://doi.org/10.1093/elt/ccq015

[41] Tannenbaum Richard and Wylie Caroline. 2008. Linking English-Language Test Scores Onto the Common European Framework of Reference: An Application of Standard-Setting Methodology. *ETS Research Report Series* 1 (2008), 1–75. https://doi.org/10.1002/j.2333-8504.2008.tb02120.x

[42] Michael Rodgers. 2013. *English Language Learning Through Viewing Television: an Investigation of Comprehension, Incidental Vocabulary Acquisition, Lexical Coverage, Attitudes, and Captions*. Ph.D. Dissertation. Victoria University of Wellington.

[43] Nathan Sakunkoo and Pattie Sakunkoo. 2013. GliFlix: Using Movie Subtitles For Language Learning. In *The Adjunct Publication of the 26th Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 7–8.

[44] Educational Testing Service. 2022. TOEFL. https://www.ets.org/toefl (Last Access Date:2022/4/6).

[45] Sunghyun Song, Jeong-ki Hong, Ian Oakley, Jun Dong Cho, and Andrea Bianchi. 2015. Automatically Adjusting the Speed of E-Learning Videos. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1451–1456. https://doi.org/10.1145/2702613.2732711

[46] Lars Stenius Staehr. 2008. Vocabulary Size and the Skills of Listening, Reading and Writing. *The Language Learning Journal* 36, 2 (2008), 139–152. https://doi.org/10.1080/09571730802389975

[47] Webb Stuart. 2015. Extensive Viewing: Language Learning through Watching Television. In *Language Learning Beyond the Classroom*. 159–168. https://doi.org/10.4324/9781315883472-24

[48] subtitlesfll.com. 2022. Subtitles for Language Learning (Prime Video). https://www.subtitlesfll.com/en.

[49] Kousuke Sugai, Shigeru Yamane, and Kazuo Kanzaki. 2016. The Time Domain Factors Affecting EFL Learners' Listening Comprehension: a Study on Japanese EFL Learners. *ARELE: Annual Review of English Language Education in Japan* 27 (2016), 97–108. https://doi.org/10.20581/arele.27.0_97

[50] Ying-li Tian, Takeo Kanade, and Jeffrey Cohn. 2001. Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 2 (2001), 97–115. https://doi.org/10.1109/34.908962

[51] Stuart Webb and Michael Rodgers. 2009. Vocabulary Demands of Television Programs. *Language Learning* 59 (2009), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

[52] Mehmet Özgen and Nazlı Gunduz. 2020. Authentic Captioned Sitcom as Listening Comprehension Material in English Language Teaching. *ELT Research Journal* 9, 2 (2020), 167–193.