English Language Learning Employing Developments in Multimedia IS

Piotr Odya
Gdansk University of Technology
Gdańsk, Poland *pioodya@pg.edu.pl*

Adam Kurowski
Gdansk University of Technology
Gdańsk, Poland *akurow@sound.eti.pg.gda.pl*

Andrzej Czyżewski
Gdansk University of Technology
Gdańsk, Poland *ac@pg.edu.pl*

Abstract

In the realm of the development of Information Systems (IS) related to education, integrating multimedia technologies offers novel ways to enhance foreign language learning. This study investigates audio-video processing methods that leverage real-time speech rate adjustment and dynamic captioning to support English language acquisition. Through a mixed-methods analysis involving participants from a language school, we explore the impact of auditory, visual, and bimodal input enhancements on learning outcomes. Results reveal that visual enhancements significantly enhance vocabulary acquisition and pronunciation, whereas simultaneous auditory-visual modifications show less advantage. Despite the limited number of participants, these findings suggest that multimedia-enhanced environments have the potential to substantially improve language learning efficiency. This research contributes to Information Systems development education by proposing practical tools and strategies for embedding multimedia content in e-learning, thereby addressing the diverse needs of students in the digital era. Further research with large sample sizes is recommended to validate these findings.

Keywords: captioning, multimedia, speech rate, time scale modification.

1. Introduction

As computer technologies continue to evolve, the field of Computer-Assisted Language Learning (CALL) is expanding its horizons, increasingly harnessing the power of multimedia content. Beyond simple exercises using graphics and text, we now have the potential to fully exploit film materials and sound recordings. Moreover, real-time content adjustments using digital signal processing algorithms are now within reach. These advancements not only make language teaching more engaging, but they also hold great promise for Information Systems Development (ISD) education. Understanding and implementing such technologies can significantly enhance the design and development of educational systems, paving the way for a more effective and engaging learning experience.

1.1. Rate of speech

The motivation for our research stems from experiments with individuals with Central Auditory Processing Disorders (CAPD), who typically face challenges in understanding speech in difficult listening environments. This study addresses these individuals' needs

and explores how digital signal processing techniques can be applied in ISD education to improve speech intelligibility, thereby enhancing language learning tools and methodologies [14].

The recommendations for CAPD refer to strategies for improving the acoustic environment, compensation, and training methods. One of the solutions is methods based on Time Scale Modification algorithms (TSM). The principle of their operation is to extend the speech duration without changing the material pitch. These methods are based on the assumption that the additional time obtained by slowing down the speech allows people with CAPD to assimilate better the information reaching them, which may lead to improved speech understanding.

In a more general context, it is worth noting that the problem with comprehension of fast speech is not limited only to people with CAPD (and other similar disorders). Even healthy people learning a foreign language may experience difficulties in understanding utterances produced with fast speech rates. It should also be noted that the ability to comprehend an utterance is mandatory for communicating with someone who speaks a foreign language. Moreover, an L2 listener entirely depends on the speed the interlocutor produces. While readers perceive separated words and can pause to rethink the meaning of phrases, listeners must immediately process the speech in real-time and divide it into different parts. It makes speech partitioning extremely difficult [10], [16].

Positive effects of lower speech rates were observed by Blau [4], Jensen & Vinther [17], and Griffith [12]. However, Blau also noticed the adverse effects of slowing down speech in the case of students with a higher proficiency level. Le also reported difficulties unambiguously assessing the obtained participants' scores [21]. On the contrary, Hayati's research shows that participants who listened to the natural speech rate achieved better scores in listening comprehension than those who used slow speech versions [15]. Similar effects were observed by Bowles and Healy [5] and Griffith [13].

Interesting observations regarding the efficiency of slowing down speech in L2 learning have been made by Zhao [30]. He pointed out that the main problem concerning other experiments is the inability to set the recording speech rate according to participants' needs. He proved that the participants obtained better scores when they could control the speech rate and then hear the recording again.

Unfortunately, the inconsistency and inconclusiveness of the obtained results are noticeable while reading literature regarding the effectiveness of teaching L2 students by reducing the speech rate. Results are affected by numerous parameters related, among others, to the students' proficiency, the definition of "fast" or "slow" speech rates, and the speech rate modification method. The latter factor is crucial since it may affect speech quality. Some researchers used natural speech (articulated faster or slower [15], [17]), but most experiments employed processed speech. In the 1990s, researchers usually played the tape at a reduced speed [4], while later, they used additional sound editing software and finally employed real-time sound processing algorithms. The topic of selecting the best algorithms for learning foreign languages can also be found in the literature. For example, Demol et al. [7] and Donnellan et al. [8] discuss existing Time Scale Modification of speech algorithms and propose new ones. Kupryjanow and Czyzewski proposed an algorithm to improve speech perception [18]. According to the results documented by the authors, the algorithm ensures the high quality and naturalness of the subjectively perceived speech.

1.2. Captions

While altering speech rate does not constantly improve the performance of L2 learners, the results of learning supported by captions are promising [6]. This distinction between subtitles (different languages) and captions (transcriptions) is crucial for developing practical multimedia learning tools in ISD education. ISD students can create more effective and engaging language learning systems by understanding and integrating such techniques.

The first observations indicating the usefulness of captions date back to the mid-1980s when captions accompanied analog TV broadcasts. More recently, the topic was studied

by Bensalem [3] and Teng [27]. In general, groups with captions obtained much better results than the others. In addition, Teng proposed the idea of using highlighted keywords and captions with glossed keywords. In turn, Sydorenko [25] showed that different groups acquired other skills: groups with captions obtained better results in learning written forms of vocabulary than groups without captions – in learning aural forms. Researchers also verified discrepancies between captions and transcriptions. Intermediate L2 learners participated in the test provided by Grgurović and Hegelheimer [11]. The participants preferred using captions to transcriptions, although both groups obtained similar results.

Some drawbacks of captions can also be found in the literature. For example, Yeldham exposed that the less proficient learners have problems since they usually read texts and do not listen to the speakers [29]. [1]. His observations confirm that low-proficiency learners achieve better results when they watch videos (vlogs) without accompanying captions. In contrast, high-proficiency learners gain higher listening comprehension scores with captions.

2. Material and methods

Building on the existing knowledge, we conducted this study to determine whether decreasing the captioned speech rate and highlighting words could improve the understanding and memorization of audiovisual information for L2 learners. This approach has significant implications for ISD education, where the development of advanced multimedia learning applications can benefit from such findings, providing students with practical tools and strategies for embedding multimedia content in e-learning environments.

The study used a specially developed multimedia application based on a stimulator created for the needs of children with lateralization disorders described in our earlier paper [20]. In this paper, we discuss its use by English language learners. The main idea behind the application is to perform a simultaneous auditory and visual input enhancement using digital signal processing techniques. The block diagram of the application is presented in Figure 1. The application processes sound files using the developed speech rate modification algorithm and displays corresponding video and captions files synchronously to the processed sound. The functionalities of the application is described in Section 2.1.

We decided to use speech rate reduction, bearing in mind that the literature review results suggest that it can be a suboptimal approach to solving such a language-learning problem. First, however, we wanted to investigate if the proposed speech stretching algorithm had some features that differentiated it from those used in studies from the literature reviews. In addition, we wanted to verify the capabilities of the algorithm in such a type of implementation because it was earlier proven to be an effective tool for improving the speech intelligibility of people with comprehension disorders.

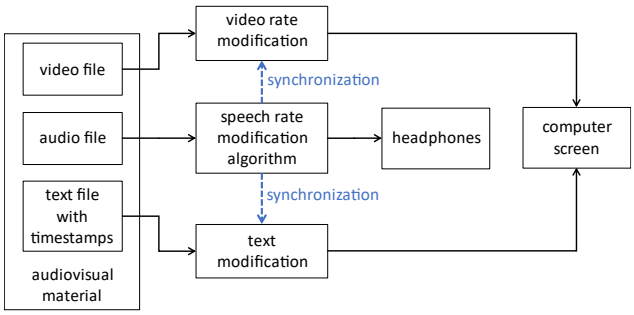

Fig. 1. Block diagram of the auditory-visual input enhancement application

The study concept and assumptions were developed by an L2 expert with more than 15 years of experience developing concepts and deploying multimedia language courses (mainly English language courses). We wanted to confirm the presence of:

- the effect of joint auditory-visual input enhancement (achieved through slowing down the speech and highlighting the captions together),

- the effect of exclusive auditory input enhancement (speech slow down only),
- the effect of exclusive visual input enhancement (synchronous captions highlighting only).

Moreover, we wanted to compare the effectiveness of English language learning with and without auditory-visual input enhancement in the following skills:

- understanding the meaning of speech - an essential skill for using a foreign language,
- assimilation of new contextual vocabulary related to the utterance - an essential element of broadening the scope of vocabulary and self-education necessary to achieve good communication skills,
- pronunciation skills - key language skills, vital in lowering the anxiety levels of a learner defined in the literature [26],
- the ability to spell correctly.

2.1. Auditory-visual input enhancement

Enhancement of the auditory input relies on modification of the speech duration. It employs the algorithm proposed by Kupryjanow and Czyzewski. A detailed explanation of how the algorithm works can be found in earlier papers [18] and [19], therefore in this article, it will be discussed only briefly.

The algorithm primary assumption is to modify the duration of different speech units using various time scaling factors in real-time. First, the speech signal is analyzed, whereby vowels, consonants, and pauses are detected and marked for further processing. The speech duration can be modeled in numerous ways, i.e., vowels can be stretched and consonants not. In our study, vowels were stretched using higher scale factor values than consonants. This time-expansion strategy is similar to the typical human behavior when a person speaks slower (speakers tend to stretch vowels more than consonants) [9]. Pauses were stretched without additional processing. The core of the TSM method is the SOLA (synchronous overlap-add) algorithm [23]. It provides high-quality, slowed-down speech and is not computationally demanding [7].

The application allows for the simultaneous transposing of the speech signal to each ear. Still, in our study, the signal for both ears was equally processed (Figure 2). This option may be helpful for people with auditory lateralization disorders. The same applies to the option "Sync to Stretched Speech." If unchecked, the words highlighted are synchronized to the original, unprocessed sound.

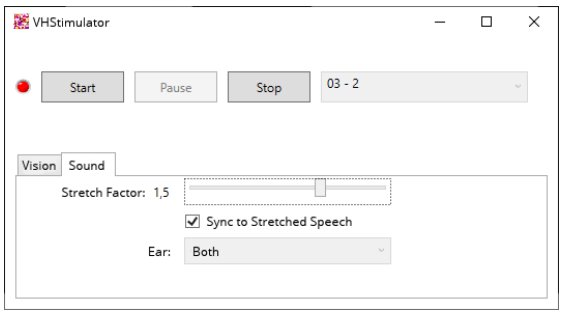

Fig. 2. Auditory input enhancement control panel

The visual input enhancement control panel (Figure 3) has only one option: the font face. A sans-serif font (Arial) was used to obtain better readability of the captions. All other options, such as plain or highlighted text color, can be set independently in the configuration file (in XML format). It is also possible to switch the highlighting off. This option was used for the participants from the group with auditory-only input enhancement and the control group. For each audio-only file or audio file (from an audio-video file), a corresponding text file containing captions (with precise timestamps of each word or sentence) was prepared.

The study plain text color was set to black, and the background was white. The red font was used as a text highlighter for the sentence parts for the participants from the bimodal

and visual input enhancement groups. The aim was to focus the participant's attention on defined parts of the speech (e.g., phrases used in the vocabulary activity). Thus, the highlighting was always synchronized with the transposed speech.

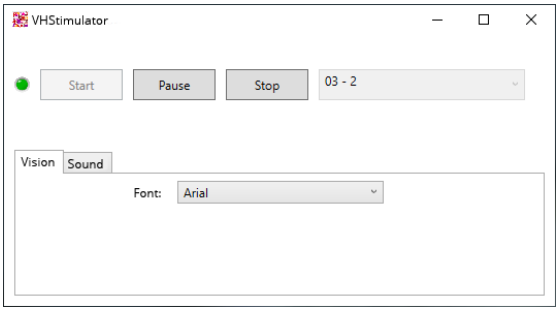

Fig. 3. Visual input enhancement control panel

The text location depended on the file type. In audio-video files, the text window was located below the presented video (Figure 4a). Contrary to typical subtitles, the captions were not overlaid on the video frame. Captions presented in this form helped learners comprehend a larger part of the text without impacting the presented video. For audio-only files, the text was located in the upper part of the screen (Figure 4b).

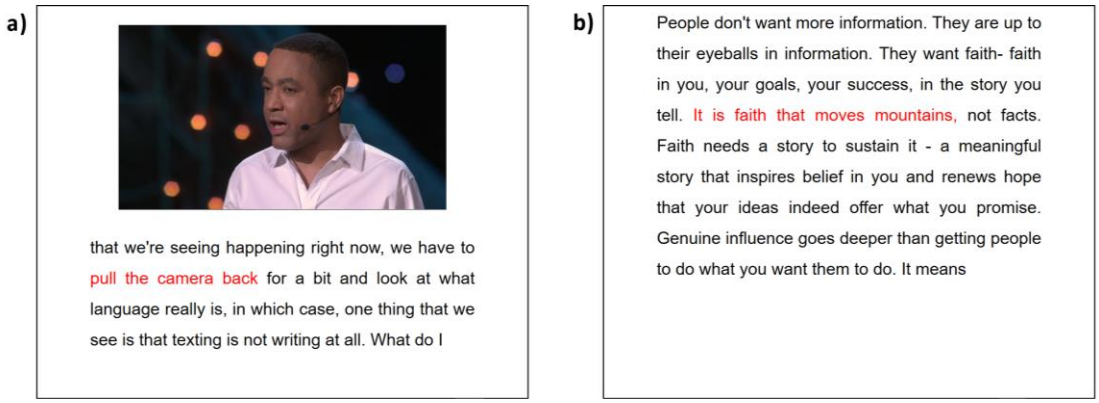

Fig. 4. Image captured from one of the captioned videos (a) and one of the captioned audio files (b) used for the study. The highlighted phrase (in red font) is visible (video from the www.ted.com website).

2.2. Material

The audiovisual materials took the form of a presentation led by a speaker to the audience (e.g., from the TED website [28]) or a 'radio' program discussing a phenomenon or event. The materials always contained key thoughts and specialized vocabulary that should be learned. The exercises assessed the perception of these thoughts.

Each participant used a dedicated computer station on which the audiovisual material was installed, plus a developed application (described in Section 2.1). The participants listened to the audio material through headphones. Each participant received a printed lesson questionnaire – they filled it in with responses and answers according to the lesson progress. After each lesson, the teacher scored the results according to the answer key. In addition, the teacher assessed the pronunciation exercises during the classes. The participants received scores in points.

The participants could pause and resume a played file. All recordings were played twice during the same exercise to help participants extract the essential parts of the material and assimilate the correct pronunciation. The stretching rate was set to 1.5 during the first viewing/listening and 1.25 during the second for the bimodal and auditory input enhancement groups. The participants from the other two groups listened to the material twice without changing signal processing parameters (the stretching rate was set to 1).

Each participant had to complete 139 exercises organized into 15 similarly organized

lessons. The lesson content was the same for all the groups. A typical lesson consisted of 6 types of activities:

- text comprehension (comprehension) – up to 3 exercises (true/false or multiple choice),
- remembering the use of new words (vocabulary) – up to 3 exercises (gap-filling),
- remembering meanings of words from the previous lesson of the study (vocabulary-recall) – 1 exercise (writing a translation of words),
- spelling correctness when writing words by ear (spelling) – 1 exercise,
- pronunciation of words learned during the given lesson (pronunciation),
- remembering the pronunciation of words from the previous study lesson (pronunciation-recall).

Both activities concerning pronunciation were combined to form one exercise since it was more convenient for the teacher who supervised the lessons.

2.3. Participants

The study was conducted at a private language school under the teachers' supervision. However, the teachers did not plan the research and its objectives. The participants were school students who responded to information about the possibility of taking part in the study. All participants were native Polish speakers who declared their knowledge of English at the intermediate level.

The study involved 40 adults and lasted three weeks. The participants were divided randomly into four 10-person groups. Unfortunately, several participants did not complete the study (have decided not to participate in the survey). Therefore, in total 32 people completed all lessons:

- the group with auditory and visual (bimodal) input enhancement – 9 participants,
- the group with auditory-only input enhancement – 8 participants,
- the group with visual-only input enhancement – 6 participants,
- the control group (without any input enhancement) – 9 participants.

3. Results and analysis

The score for each exercise was normalized using the following formula:

$$\text{score}_{\text{norm}} = \frac{\text{points}_{\text{acq}}}{\text{points}_{\text{total}}} \tag{1}$$

where:

- $\text{score}_{\text{norm}}$ is the normalized score gained by a participant,
- $\text{points}_{\text{acq}}$ is the number of points acquired by a participant in a given exercise,
- $\text{points}_{\text{total}}$ is the maximum possible score achievable in the exercise.

Thus, after normalization, the number of points scored in the exercise is confirmed in the range [0,1].

The dataset comprises 4448 data points (32 participants x 139 activities). Each data point corresponds with a result of a single activity performed by a given person during a lesson.

The analysis based on mixed linear models is widely used in research fields such as psychology, linguistics, and medical research [22], [24]. We employed linear mixed-effects models to analyze the repeated measurement results of the participants' performance across a set of selected linguistic skills. These models are particularly suited for handling datasets with repeated measures and missing values. The missing values in our case were due to the variable count of repetitions of exercises throughout the study. The rationale for using linear mixed-effects models is as follows:

- mixed-effects models can account for the correlations between repeated measurements taken from the same participants;
- they allow for the inclusion of all available data without excluding participants with incomplete data;
- by not averaging observations, we preserve the variability and avoid eliminating the influence of confounding variables such as the number of lessons or the type of exercises.

The study included two main factors:

1. Confounding variables: variables like lesson ID and the repetition of skill-related activities during each lesson.
2. Random factors include the lesson ID and skill repetitions, which were modeled as random effects.

The dependent variable in our study was the normalized score. The main goal was to find how the values of the normalized scores varied across different groups, so we considered two explanatory variables:

1. Input enhancement type: the type of input enhancement (auditory, visual, or bimodal).
2. Skill type: the specific skill to which the exercises were connected.

The potential influence of input enhancement type and lesson ID on the normalized score was also considered. Different groups of participants were exposed to different input enhancements, and their linguistic skills varied significantly. We accounted for this variability by using a random slope variant of the mixed-effects linear model, which considers the effect of different starting points for each group of participants.

3.1. Statistical model definitions

The analyses were performed using the lme4 package [2] in R with the following model definitions:

1. Full model:

$$normalised\_score \sim input\_enhancement\_type + skill + (1 + input\_enhancement\_type | lesson\_id) + (1 | skill\_repetitions) \quad (2)$$

where:

- dependent variable: *normalized_score*
- explanatory variables: *input_enhancement_type*, *skill*
- random effects:
  - the influence of lesson ID on the effectiveness of the input enhancement type (more pronounced for later lessons).
  - the influence of skill-related exercise repetitions during the training lesson.

2. Null hypothesis model for likelihood ratio test:

$$normalized\_score \sim skill + (1 + input\_enhancement\_type | lesson\_id) + (1 | skill\_repetitions) \quad (3)$$

The likelihood ratio test comparing this null model to the full model yielded a χ^2 test statistic value of 24.136 with a p-value of less than 0.001, indicating that the differences between models are statistically significant.

3. Skill-related dataset with skill repetitions:

$$normalized\_score \sim input\_enhancement\_type + (1 + input\_enhancement\_type | lesson\_id) + (1 | skill\_repetitions) \quad (4)$$

4. Skill-related dataset without skill repetitions:

$$normalized\_score \sim input\_enhancement\_type + (1 + input\_enhancement\_type | lesson\_id) \quad (5)$$

3.2. Analysis of the set containing all skills

The results of analyses performed with mixed-effects linear models obtained for the dataset containing all skills are presented in Table 1. The skill is treated here as one of two explanatory variables. Each row describes how the type of skill affects the performance of participants. Confidence intervals were calculated in the R programming language by employing the confint function with Wald's method specified for calculation. Underlining denotes improving participants' performance, and the grey background indicates the observed performance degradation. Statistically significant values of explanatory variables are marked with a bold italic font.

Table 1 contains a column depicting the explanatory variable name and its categorical values to which each row of the table is assigned. Next are columns containing the mean values and standard deviations of each explanatory variable expected values on the dependent variable mean (normalized score). The first row of the table includes a particular value called the intercept. It is an intercept point of the linear model that can be interpreted

as a reference point for all subsequent rows. In our study, the intercept is the mean grade the control group participants obtained for the text comprehension skill. All other rows show the relative difference in the participants' performance measured by a normalized score. It is always expressed as a value that must be added to the reference intercept row to obtain the values associated with a given value of the explaining variables. The t column contains the value calculated by dividing the mean estimate of the normalized score by a standard deviation. It can be used to assess the strength of the effect imposed by each explanatory variable on the normalized score. The last two columns contain the left and right boundary of the confidence interval in which the mean of the intercept of the influence of the explanatory variables can be found. The significance level for the confidence interval is set to a standard value of 0.05.

Table 1. Results of the mixed-effects model analysis applied to the dataset containing all skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group + comprehension skill)	***0.772***	***0.016***	***49.220***	***0.741***	***0.803***
Input enhancement: auditory	0.000	0.009	0.050	-0.017	0.017
Input enhancement: visual	***0.051***	***0.011***	***4.460***	***0.029***	***0.073***
Input enhancement: bimodal	-0.015	0.009	-1.770	-0.033	0.002
Skill: pronounce	***-0.024***	***0.012***	***-1.950***	***-0.048***	***0.000***
Skill: pronounce - recall	0.011	0.015	0.780	-0.017	0.040
Skill: recall - vocabulary	***-0.394***	***0.012***	***-31.660***	***-0.418***	***-0.369***
Skill: spelling	0.020	0.012	1.680	-0.003	0.044
Skill: vocabulary	***0.142***	***0.008***	***17.590***	***0.126***	***0.157***

The results are interpreted as statistically significant if the confidence interval containing the relative change value of the participants' performance does not include zero. Such a situation can be observed for the visual input enhancement, improving the participants' performance, pronounce recall-vocabulary, and vocabulary skills.

The analysis of the dataset containing all skills showed the following key points:

- visual input enhancement: statistically significant improvement in participants' performance regarding pronunciation recall vocabulary and vocabulary skills. Confidence interval for improvement: [0.029, 0.073], indicating that the visual input enhancement group gained roughly 3 to 7 points more than the control group (the maximum number of points to be scored was 100).
- auditory input enhancement: no significant improvement or degradation in overall performance.
- bimodal input enhancement: slight performance degradation, particularly in text comprehension tasks, suggesting that simultaneous auditory and visual stimuli might overwhelm participants.

Further analysis of Table I also leads to the conclusion that participants were more likely to make progress in terms of gained scores in vocabulary-related activities than in comprehension skill reference, as their normalized score, in this case, is higher by a value in a range of [0.126; 0.157]. On the other hand, the recall-vocabulary exercises were more likely to result in worse scores than the comprehension skill reference, and the deterioration in the score gained by participants is between values of 0.369 and 0.418.

3.3 Influence of input enhancement on tasks related to specific skills

To further investigate the influence of auditory-visual input enhancement on participants' performance, a more precise analysis must be completed in a per-skill manner to determine which skills are affected by the input enhancement and which skills are not. This was obtained by selecting data points related to scores obtained only from exercises carried out for just one skill type and then performing an analysis employing a mixed-effects model separately for each subset derived in such a manner.

In the case of text comprehension skills (Table 2), no statistically significant effect was found for the enhancement of auditory and visual input. Bimodal input enhancement

caused statistically significant performance degradation, suggesting long-term research, as reading comprehension skills require longer work on the language material to see improvement. Lack of change or regression may also be due to increased recording duration in recent lessons, causing the participants to lose attention.

However, each confidence interval associated with input enhancement types as a boundary is close to zero. Therefore, it is theoretically possible in such cases that the input enhancement influences participants' performance in a way that is so small that our statistical model wasn't able to detect such influence, especially if compared to the strength of the effects observed for other skills.

Table 2. Mixed-effects model analysis applied to the dataset containing text comprehension skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.777***	***0.022***	***36.030***	***0.735***	***0.819***
Input enhancement: auditory	0.017	0.016	1.060	-0.015	0.049
Input enhancement: visual	0.035	0.020	1.730	-0.005	0.074
Input enhancement: bimodal	***-0.038***	***0.017***	***-2.260***	***-0.071***	***-0.005***

For the vocabulary-related skills (Table 3), both auditory and visual input enhancements showed statistically significant improvements, with the visual enhancement having a slightly higher positive impact.

The lower boundary of the confidence interval corresponding to the visual input enhancement is higher than for the auditory one. However, both confidence intervals overlap. Therefore, the influence should be considered similar in both cases. The bimodal input enhancement confidence interval contains mostly positive values; however, its impact is considerably lower than the unimodal ones. The use of the auditory input enhancement may cause an increase in performance expressed in terms of confidence interval ranges from 0.033 to 0.078. The visual input enhancement may increase performance by widening the confidence interval from 0.048 to 0.101. However, bimodal input enhancement again resulted in the lowest performance. Participants seem to be overwhelmed by two simultaneous stimuli.

Table 3. Mixed-effects model analysis applied to the dataset containing vocabulary use and memorization skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.882***	***0.012***	***75.760***	***0.860***	***0.905***
Input enhancement: auditory	***0.056***	***0.012***	***4.770***	***0.033***	***0.078***
Input enhancement: visual	***0.075***	***0.014***	***5.520***	***0.048***	***0.101***
Input enhancement: bimodal	0.021	0.012	1.790	-0.002	0.044

No statistically significant influence was observed on spelling skills (Table 4). The visual input enhancement showed potential improvement, although the results were not statistically conclusive.

It is also the only input enhancement type that, in the worst case, will not negatively affect performance. Such results, or even better, were expected - highlighting a word (or a sequence of words) might significantly aid the memorization process.

Table 4. Mixed-effects model analysis applied to the dataset containing spelling skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.784***	***0.035***	***22.550***	***0.716***	***0.852***
Input enhancement: auditory	0.036	0.028	1.300	-0.018	0.091
Input enhancement: visual	0.062	0.032	1.950	0.000	0.124
Input enhancement: bimodal	0.021	0.027	0.770	-0.032	0.074

Regarding pronunciation skills (Table 5), the visual input enhancement significantly improved performance, while the auditory input enhancement might have been detrimental due to the chosen speech slowdown rates. Perhaps the rates used were too high and negatively affected the process of remembering the correct pronunciation.

The confidence interval for the visual input enhancement in the case of the pronunciation skill denotes an increase in participants' performance, with a value in a confidence interval ranging from 0.059 to 0.145. Therefore, the use of visual input enhancement is recommended for pronunciation-related exercises.

Table 5. Mixed-effects model analysis applied to the dataset containing pronunciation skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.751***	***0.024***	***31.650***	***0.705***	***0.798***
Input enhancement: auditory	-0.033	0.021	-1.590	-0.073	0.008
Input enhancement: visual	***0.102***	***0.022***	***4.670***	***0.059***	***0.145***
Input enhancement: bimodal	0.014	0.019	0.740	-0.023	0.051

For vocabulary-recall exercises (Table 6), the auditory and bimodal input enhancement caused degradation in performance, indicating that while these enhancements aid initial learning, they might negatively impact long-term recall. More research is needed to find precisely the reason for such a phenomenon.

There was no statistically significant degradation in the case of visual input enhancement. However, the confidence interval for this case contains mainly negative values.

Table 6. Mixed-effects model analysis applied to the dataset containing vocabulary-recall skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.486***	***0.042***	***11.610***	***0.404***	***0.568***
Input enhancement: auditory	***-0.181***	***0.034***	***-5.280***	***-0.248***	***-0.114***
Input enhancement: visual	-0.046	0.038	-1.220	-0.119	0.028
Input enhancement: bimodal	***-0.109***	***0.034***	***-3.210***	***-0.176***	***-0.043***

For the pronunciation-recall skills (Table 7), the auditory input enhancement negatively affected recall performance (similar to the vocabulary-recall skills). In such cases, it is necessary to learn, i.e., how to pronounce a given word at the moment and recall it after taking part in the previous lesson. We hypothesize that this additional input enhancement may harm this recall-related part of the exercise. In this case, it should also be noted that the right boundary of the confidence interval is close to zero. Therefore, according to our results, the visual analysis may lead to a considerable increase in performance (0.119 in terms of normalized score). However, it can slightly degrade performance (-0.002 in normalized score).

Table 7. Mixed-effects model analysis applied to the dataset containing pronunciation-recall skills data.

Explanatory variable value	norm. score estimate	std. dev.	t	left cfb	right cfb
Intercept (control group)	***0.836***	***0.033***	***25.530***	***0.772***	***0.900***
Input enhancement: auditory	***-0.058***	***0.025***	***-2.330***	***-0.107***	***-0.009***
Input enhancement: visual	0.059	0.031	1.910	-0.002	0.119
Input enhancement: bimodal	-0.040	0.024	-1.640	-0.087	0.008

4. Discussion and conclusions

This study has delved into the intersection of Information Systems (IS) Development Education and multimedia learning technologies, spotlighting the impact of innovative

digital content processing on enhancing English language learning. The imperative drove our exploration to integrate cutting-edge educational tools into the curriculum for IS specialists, aligning with the broader themes of creativity, innovation, instructional design, and Human-Computer Interaction (HCI) in education.

Our findings show the significant potential of multimedia content enhancements – specifically, real-time speech rate reduction and dynamic captioning – to foster more effective and engaging learning experiences. This aligns with the emphasis on computer-supported collaborative learning, as the technologies we have developed and assessed offer new pathways for collaborative and interactive learning environments. These multimedia tools can be seamlessly integrated into IS development education, providing a model for incorporating audiovisual materials into complex subject matter teaching.

The innovative application of audio and video processing applications presented in this research underscores the creativity and innovation critical to IT-based education. Our work goes beyond conventional language learning tools by providing a multimedia learning environment that supports various learner needs through customized content presentation. This innovation in educational technology design and implementation exemplifies the transformative potential of integrating sophisticated digital processing techniques into educational settings.

However, it is essential to note that the limited number of participants in this study may affect the generalizability of the results. The small sample size may limit the statistical power of the findings, making it necessary to interpret the results cautiously. Despite this limitation, the observed trends provide valuable insights that warrant further investigation. Future research with more extensive and diverse samples is essential to confirm the validity and reliability of these results and explore additional variables that might influence the effectiveness of multimedia content enhancements.

Our research also touches on HCI issues in IS development for education, particularly how intuitive design and user interaction with multimedia content can enhance the learning process. The positive learning outcomes observed in our study highlight the importance of user-friendly educational technologies that cater to diverse learning preferences and needs. This focus is especially pertinent to IS development education, where understanding and applying HCI principles can significantly influence the effectiveness of educational software and systems.

In conclusion, our investigation into multimedia content enhancements for English language learning provides compelling evidence of the benefits of such technologies in enhancing the effectiveness and appeal of educational offerings. We aimed to contribute to the ongoing dialogue on the future of IS development education. We have argued for leveraging innovative technologies to create inclusive, engaging, and effective learning environments. It may not only enrich the educational experiences of future IS specialists but also equip them with the skills and knowledge necessary to thrive in a digitally interconnected world.

References

1. Aldukhayel, D.: The effects of captions on L2 learners' comprehension of vlogs. Language Learning & Technology ISSN. 25 (2), 178–191 (2021)
2. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 67 (1), 1–48 (2015)
3. Bensalem, E.: The Efficacy of Captions on Students' Incidental Vocabulary Acquisition. Journal of Teaching and Teacher Education. 6 (1), 1–11 (2018)
4. Blau, E.K.: The Effect of Syntax, Speed, and Pauses on Listening Comprehension. TESOL Quarterly. 24 (4), 746–753 (1990)
5. Bowles, A.R., Healy, A.F.: Training and Transfer of Word Identification: Foreign Language Speech Rate. J Appl Res Mem Cogn. 6 (3), 253–259 (2017)
6. Danan, M.: Captioning and Subtitling: Undervalued Language Learning Strategies. Meta. 49 (1), 67–77 (2004)
7. Demol, M., Verhelst, W., Struyve, K., Verhoeve, P.: Efficient Non-uniform Time-scaling Of Speech With WSOLA. In: SPECOM 2005. pp. 163–166. SPECOM 2005, Vol. 1, pp.

163-166, Patras, Greece., Patras, Greece (2005)
8. Donnellan, O., Jung, E., Coyle, E.: Speech-adaptive time-scale modification for computer assisted language-learning. In: Proceedings 3rd IEEE International Conference on Advanced Technologies. pp. 165–169. IEEE Comput. Soc, Athens, Greece (2003)
9. Ebihara, T., Ishikawa, Y., Kisuki, Y., Sakamoto, T., Hase, T.: Speech synthesis software with variable speaking rate and its implementation on a 32-bit microprocessor. In: 2000 Digest of Technical Papers. International Conference on Consumer Electronics. Nineteenth in the Series (Cat. No.00CH37102). pp. 254–255. IEEE, Los Angeles, CA, USA (2000)
10. Graham, S.: Listening comprehension: The learners' perspective. System. 34 (2), 165–182 (2006)
11. Grgurović, M., Hegelheimer, V.: Help options and multimedia listening: Students' use of subtitles and the transcript. Language Learning & Technology. 11 (1), 45–66 (2007)
12. Griffiths, R.: Speech Rate and Listening Comprehension: Further Evidence of the Relationship. TESOL Quarterly. 26 (2), 385 (1992)
13. Griffiths, R.: Speech Rate and NNS Comprehension: A Preliminary Study in Time-Benefit Analysis. Lang Learn. 40 (3), 311–336 (1990)
14. Guiraud, H., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Caillot-Bascoul, A., Gonzalez-Monge, S., Boulenger, V.: Don't speak too fast! Processing of fast rate speech in children with specific language impairment. PLoS One. 13 (1), e0191808 (2018)
15. Hayati, A.: The Effect of Speech Rate on Listening Comprehension of EFL learners. Creat Educ. 1 (2), 107–114 (2010)
16. Hulstijn, J.H.: Connectionist Models of Language Processing and the Training of Listening Skills With the Aid of Multimedia Software. Comput Assist Lang Learn. 16 (5), 413–425 (2003)
17. Jensen, E.D., Vinther, T.: Exact Repetition as Input Enhancement in Second Language Acquisition. Lang Learn. 53 (3), 373–428 (2003)
18. Kupryjanow, A., Czyzewski, A.: A Method of Real-Time Non-uniform Speech Stretching. In: Obaidat, M.S., Sevillano, J.L., and Filipe, J. (eds.) E-Business and Telecommunications. pp. 362–373. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
19. Kupryjanow, A., Czyzewski, A.: Methods of Improving Speech Intelligibility for Listeners with Hearing Resolution Deficit. Diagn Pathol. 7 (1), 129 (2012)
20. Kupryjanow, A., Kosikowski, L., Odya, P., Czyzewski, A.: Auditory-Visual Attention Stimulator. In: Proceeding of the 134th Audio Engineering Society Convention. , Rome, Italy (2013)
21. Le, F.: Faster, normal or slower?: the effects of speech rates on high-intermediate ESL learners' listening comprehension of academic lectures. Iowa State University, Digital Repository (2006)
22. Magezi, D.A.: Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). Front Psychol. 6 (2015)
23. Roucos, S., Wilgus, A.M.: High quality time-scale modification for speech. In: ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing. pp. 493–496. Institute of Electrical and Electronics Engineers, Tampa, FL, USA (1985)
24. Speelman, D., Heylen, K., Geeraerts, D. eds: Mixed-Effects Regression Models in Linguistics. Springer International Publishing, Cham (2018)
25. Sydorenko, T.: Modality of input and vocabulary acquisition. 14 (2), 50–73 (2010)
26. Szyszka, M.: Pronunciation Learning Strategies and Language Anxiety. Springer International Publishing, Cham (2017)
27. Teng, F. (Mark): Vocabulary learning through videos: captions, advance-organizer strategy, and their combination. Comput Assist Lang Learn. 35 (3), 518–550 (2022)
28. Wingrove, P.: How suitable are TED talks for academic listening? J Engl Acad Purp. 30 79–95 (2017)
29. Yeldham, M.: Viewing L2 captioned videos: what's in it for the listener? Comput Assist Lang Learn. 31 (4), 367–389 (2018)
30. Zhao, Y.: The Effects of Listeners' Control of Speech Rate on Second Language Comprehension. Appl Linguist. 18 (1), 49–68 (1997)