

Application of Perceptual Model-Driven Multimedia Information Retrieval Technology in English Teaching Management Systems

Jingfang Sun, Shandong Foreign Trade Vocational College, China*

ABSTRACT

The application of multimedia technology in college English classroom teaching is suitable for sports and mute with illustrations, excellent audio, and video. The application of multimedia in the field of education has become the focus of the current English education reform, and education has entered the industrial age and the information age. This research aims at teaching practice in order to obtain high-quality educational results. And gradually turn multimedia, a new way, into an effective teaching tool in English education. Based on the research on the demand of information perception for English teaching, this article studies the relationship between constant information perception and English teaching demand from the perspective of multimedia information perception. An evolving model of English teaching needs caused by the change of multimedia information perception due to the input of external resources is established. The study found that when the average consumer perception is low, the revenue of enterprises can be rapidly improved by investing external resources.

KEYWORDS

English education, Information retrieval, management system, Multimedia technology

INTRODUCTION

Learning style is a habitual learning method with individual characteristics, which is the sum of learning strategies and learning tendencies. The theory of perceptual learning style has been deeply investigated and found that learners acquire learning experience through different senses. The breakthrough in information and communication technology (ICT) has brought new opportunities for restructuring the language learning/teaching environment. Nowadays, information and communication technology has opened up new ways and brought new challenges for language learners and teachers. The responsibility of learning tasks has shifted to learners, which has greatly changed the role of teachers (Benavent et al., 2013). Everyone has his own preferred learning perception channel or unique learning style. Perceived learning style is an important theory to introduce learners' individual learning differences. It is a perceptual channel for learners to organize, understand, and remember

DOI: 10.4018/IJWLTT.338718

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

their learning experiences. Only through this way can learners perceive and internalize the knowledge they have learned more effectively if they want to receive various kinds of information through multi-channel and multisensory means. Based on this theoretical research, a group of perceptual learning style preference questionnaires were designed. In the questionnaire, the learning styles were divided into the following types: visual, tactile, group, auditory, kinesthetic, and personal. The research on learning style and teaching shows that whether learning style and teaching match directly affects the improvement of students' learning efficiency. The research found that "teaching based on the learning style theory can improve students' academic performance" (Yu & Brandenburg, 2011). To achieve the best learning effect, students should adapt to their learning style and give full play to their learning style strengths. Students should also be able to flexibly adjust their learning style to reduce or avoid learning style deficiencies. Therefore, teachers' teaching models and strategies should be diversified and flexible.

The entry of computer information technology into the field of foreign language teaching provides material conditions and technical support for multimedia-assisted instruction and accelerates the process of educational reform (Liang, 2022). Multimedia teaching has changed the traditional teaching mode of chalk and blackboard, improved teachers' work efficiency, stimulated students' interest in learning English, and optimized English classroom teaching effects. Media can play an auxiliary role in English language teaching in senior high schools, but it is not perfect and has its own shortcomings (Asim et al., 2019). AI education (AIED) is defined as the application of AI in the field of education.

The online English teaching system developed based on artificial intelligence module and knowledge recommendation is compared with the commonly used teaching assistant system. The online English teaching system provides valuable data from a wide range of information, summarizes patterns and data, and helps teachers improve their education and students' English performance. First of all, multimedia can hold a relatively large amount of information, so that students can get richer resources and broaden their horizons. However, it is still relatively difficult to grasp the choice of key information. Secondly, although the effective application of multimedia can enhance students' interest in learning, there are often many defects in the actual courseware making process, such as the relatively beautiful interface of some courseware making (Dahl et al., 2010). Higher education institutions (HEIs) need to be more inclusive, from the methodological perspective, about what is used in the classroom. Due to the lack of available educational and teaching resources and knowledge of appropriate teaching methods and strategies, schools are unable to cultivate different language skills: listening, speaking, reading, and writing in the process of English learning. Finally, in the specific application process, because some teachers' own multimedia production technology is not skilled enough, they often blend the contents of textbooks into slides stiffly and show them to students, instead of effectively showing the advantages of multimedia. Moreover, teachers' insufficient use of multimedia technology will also lead to a waste of time and energy, which can't really improve students' learning effectiveness (Du & Liang, 2021).

The system can record the teacher's teaching situation on the teaching site, and make digital audio and video streaming media materials by using computer multimedia technology, so that students can reproduce the teacher's teaching situation and help students review after class by watching these video files during their after-school learning time (Liu et al., 2014). The goal of the application of target tracking technology in the multimedia teaching classroom recording system is to use tracking algorithms to track the captured video sequence in real time, obtain the teacher's position coordinates, and automatically adjust the direction of the PTZ camera according to the teacher's position, so that the teacher's target is within the vision of the camera, and it achieves better recording effect (Jitpaisarnwattana, 2018). The efficient and fair reflection of the teaching quality of a school not only plays a crucial role in supervising teaching but also contributes to the improvement of teaching quality through the feedback and analysis of evaluation results, as highlighted by Syakur (2020).

This perspective is particularly relevant when considering the overall phenomenon in colleges and universities, where it has been observed that students' perception of information can be subject

to manipulation. Despite this, the trend towards an increase in students' information perception is inevitable and underscores the importance of ensuring that evaluations of teaching quality are conducted in a manner that truly captures and reflects the educational experience. This connection highlights the need for mechanisms that not only accurately assess teaching quality but also safeguard against the manipulation of students' perceptions, thereby fostering an environment where the inevitable increase in information perception among students contributes positively to their educational journey. Inadequate information perception of students will lead to difficulties, and too high information perception will lead to tight teaching arrangements in colleges and universities. Therefore, colleges and universities often hope to control students' information perception to improve teaching efficiency. In this process, colleges and universities manipulate students to obtain appropriate information perception. Under the historical background that the influence of information perception on teaching demand is gradually increasing, this paper will study the change of information perception by different teaching behaviors. So as to form changes in market demand and produce analysis and research in the highly developed network environment. This study will determine the influence of information dissemination on the relationship between demand and supply of products or services in teaching and how to improve teaching requirements through external forces under goal-controlled conditions. In teaching, it is important to emphasize not only the correctness of the language, but also its fluency. Pay attention to students' emotions and attitudes, so that students can improve their independent thinking and judgment abilities and develop their communication and cooperation with others in the process of English learning. The teacher's role has changed from the classroom controller to the designer, demonstrator, instructor, and even sometimes a student (Mulyono, 2016). During the epidemic, intern teachers have been able to do online teaching activities thanks to the availability of wi-fi devices and reliable Internet connections. They are able to make the most of a wider range of applications. To transmit instructional activities, intern teachers should adjust to online learning. Students have a large space for thinking and practice in class. Through experience, participation, practice, and cooperation, they can achieve their learning goals, form a positive learning attitude, and improve their ability to use language in practice. The focus of this paper is how to design the overall teaching plan to achieve the optimization of English classroom teaching under the condition of using multimedia. It belongs to the research of teaching strategies, and the research methods are mainly literature, questionnaire, and interview. Its innovation lies in:

- (1) This paper studies the successful content-based video retrieval (CBVR) systems at home and abroad, summarizes their design methods, compares their advantages, finds their shortcomings and problems to be solved, and provides a reference and basis for the research direction and system implementation of this paper.
- (2) A video information object model is designed to facilitate the management and retrieval of video information, which provides basic data units for the system to establish a reasonable and effective video database, and also provides basic data structure types for the retrieval algorithm to process.

RELATED WORK

With the informatization of society, the information technology centered on multimedia computers and networks has been widely used in the teaching of various disciplines and has gradually affected all aspects of teaching. Foreign language teaching has also been greatly impacted, and the traditional foreign language teaching methods and teaching ideas can no longer meet the rapidly developing social environment and social requirements for talents. Teachers should help students identify types of perceptual learning styles and use scientific testing tools when necessary. To enable students to understand their own learning styles, teachers should also try to adopt teaching strategies consistent with students' perceived learning styles. As students, they should also realize the important role of

perceptual learning style in learning, understand the advantages and disadvantages of their learning style, develop their strengths and circumvent their weaknesses, learn to expand their learning style, and try to adapt to teaching methods that are inconsistent with their perceived learning style.

Schmid (2010) pointed out that “information technology should be widely used and integrated into the study of other disciplines.” Therefore, in recent years, the research on the integration of multimedia technology and ET has made great progress. With its incomparable advantages, multimedia technology has been widely used in ET, which provides new teaching models and means for ET. Chang’s (2010) research on the application of media technology in ET is on the rise. The main research objects are undergraduates, and the research contents can be summarized as follows: the significance and importance of multimedia assisted foreign language teaching, multimedia assisted foreign language teaching mode, etc. Habibi (2018) conducts research based on the cognitive principles of multimedia learning. In the experiment, it has summarized a lot of high-order rules about learning multimedia. At the same time, when many disciplines are connected with each other, it puts forward the cognitive principles of multimedia learning from a relatively microscopic perspective. Albiladi (2019) takes multimedia learning as the theme to explore the design idea and implementation of a biology interactive multimedia learning system. The system is operated based on the multimedia structure, through which multiple teaching needs, such as diversified interaction and individualized teaching, can be completed, which is of great help to improve the teaching effect. Lidas (2017) proposed the cognitive theory of multimedia learning on the basis of: dual channel theory, limited capacity hypothesis (that is, working memory capacity is limited, and the amount of information processed is also limited, if excessive, it will lead to cognitive overload), and active processing theory (learners actively process information). Brush proposes that the possible effects of increasing cognitive load can be seen by exploring learners through relevant experiments based on increasing cognitive load. The findings suggest that an increase in cognitive load cannot be equated with an increase in overall cognitive load. This strategy aims to reduce the overall cognitive load by increasing relevant cognitive load. Lauder (2010) has launched a number of communication channels, helping teachers and students to establish a good discussion space, which is very important for improving teaching quality and promoting students’ advanced cognitive ability. Buckner found that in order to improve the teaching quality, colleges and universities all over the country are constantly improving the teaching quality evaluation system and have set up teaching evaluation systems suitable for the school situation. In order to reflect the situation of multimedia network teaching in real time, the integration of multimedia interactive information network teaching and teaching evaluation system is particularly urgent. Sanprasert (2010) proposed a recursive face feature estimation method based on wavelet. In other words, the research on face tracking has only been about fifteen years. Over the years, scientists have done a lot of exploration work in this field. Alshorman (2018) adopts new machine learning technology to confirm video and audio data and add indexes to them. This system adds audio, video, and text information to multimedia content in a unique way. Designed to support MPEG-7 standard for describing multimedia content.

METHODOLOGY

In order to better combine the perceptual model-driven multimedia information retrieval technology with the English teaching management system, this paper proposes an optimization design method for English teaching based on video retrieval technology. Specifically, the method includes the following steps: first, the involved English knowledge points are transformed into corresponding teaching videos; then, these videos are categorized and retrieved using content-based video retrieval technology; during the learning process, students can quickly and accurately find the required video clips through the browsing window and various search interfaces for better understanding and mastering of English knowledge. At the same time, in order to adapt to the changes in students’ perceived demand for English knowledge, an English teaching demand evolution model of the changes in the perception of

multimedia information caused by the input of external resources will also be established, which will be used to monitor the students' perceived demand for English knowledge and to realize the dynamic adjustment of the teaching content. Through this method of optimizing the design of English teaching based on video retrieval technology, it helps to promote the innovation and development of English teaching, and improve students' learning interest and learning effect.

Overview of Multimedia Technology Combined With English Education System

At present, with the rapid development of computer technology, multimedia technology is increasingly becoming an important means of teaching modernization. The traditional teaching system consists of three elements: teachers, students, and teaching materials. In the modern teaching environment, one more element needs to be added, that is, teaching media. Multimedia teaching refers to the process of teaching, according to the characteristics of teaching objectives and teaching objects, through teaching design, reasonable selection and application of modern teaching media, and organic combination with traditional teaching methods, participating in the whole teaching process together, acting on students with various media information, forming a reasonable teaching process structure, and achieving the optimal teaching effect. Common multimedia types and characteristics are shown in Table 1.

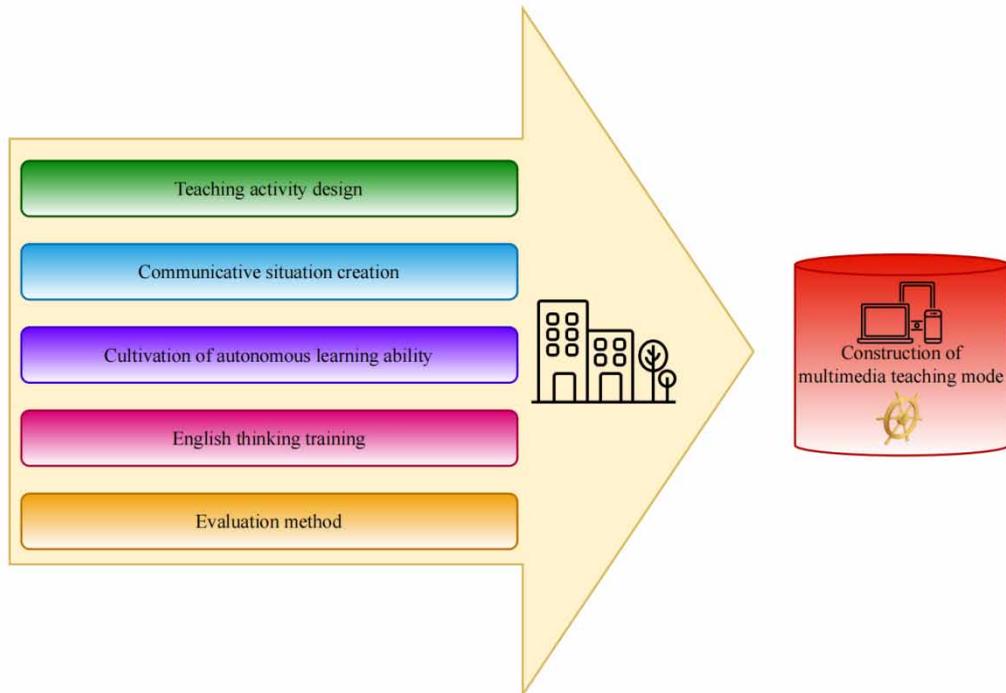
Multimedia technology includes hardware technology and software technology. Hardware technology includes storage technology, compression technology, chip technology and database technology. Software technology includes multimedia software development and multimedia software application technology. Because multimedia technology has the characteristics of integration, interactivity, and real-time, it is widely used in various fields and is imperceptibly changing people's lives. The emergence of multimedia technology also provides a new way for ET reform in higher vocational colleges (Xu & Yao, 2022). The multimedia teaching mode of New Vocational English mainly refers to the use of computer multimedia equipment, combined with the content of teaching materials, to more effectively impart language knowledge to learners through various teaching resources, such as recordings, videos, and movies, and to create a simulated vocational environment and language learning environment through multimedia technology. Based the author's experience, this paper puts the multimedia teaching mode into practice in specific teaching activities and puts forward some specific operating methods for reference and correction by peers and experts. The multimedia teaching mode is applied to specific teaching activities, and some specific operation methods are proposed for reference and correction by peers and experts. The construction of multimedia teaching mode is shown in Figure 1.

In order to achieve the goal of cultivating students' ability to use English, various means and methods should be fully used in the teaching process to mobilize students' enthusiasm for learning and maximize students' participation in the whole learning process. Under the guidance of modern teaching theory, this paper uses multimedia technology in the teaching process to analyze the content

Table 1. Media types

Medium Type	Characteristic
Feel the media	Act directly on human senses
Representing media	It is a kind of media artificially constructed for processing and transmitting sensory media, that is, various codes.
Display media	It refers to a kind of media that is converted between sensory media and electrical signals used for communication and transmission, that is, the interface between sensory media and computer. There are two types: input display media and output display media.
Storage media	Storage media: stores media representing media.
Transmission medium	The physical carrier of transmission, that is, the physical carrier used to transmit media from one place to another.

Figure 1. Construction of multimedia teaching mode



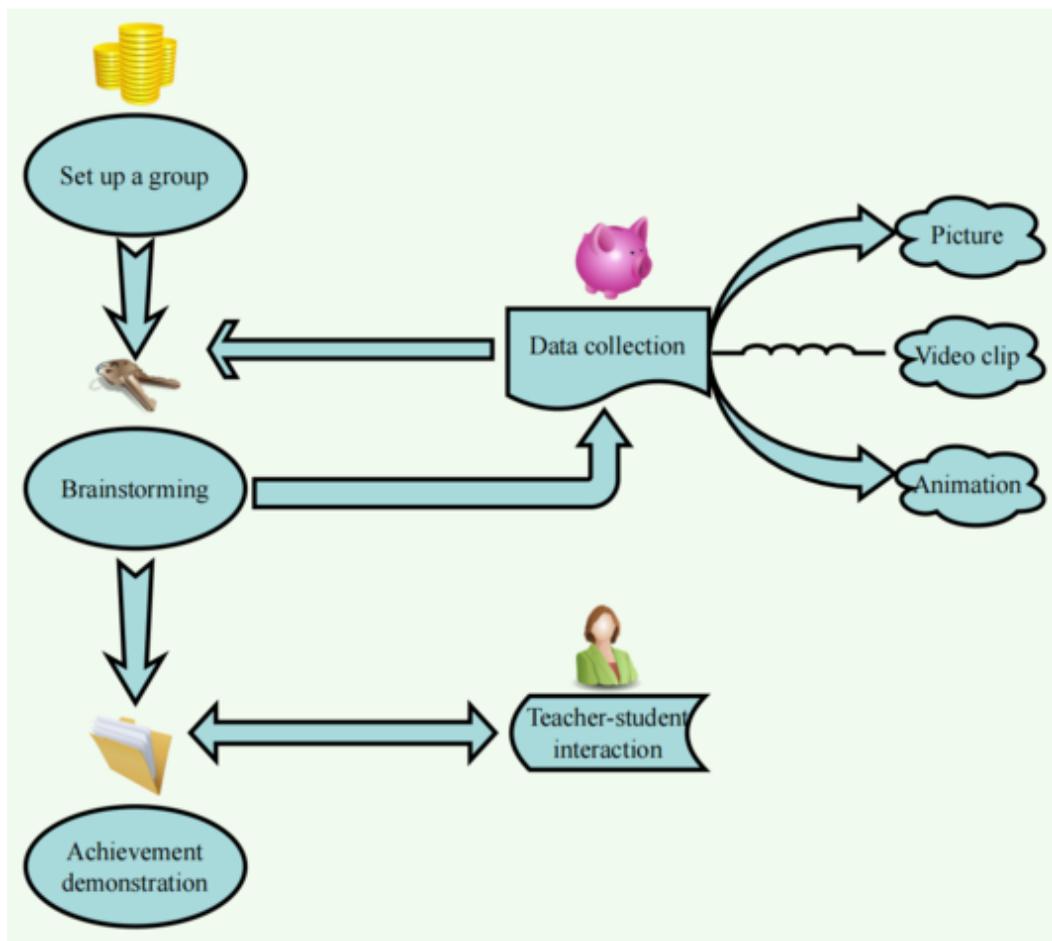
of the textbook when preparing lessons, make courseware and present it in the form of diagrams, pictures, or animations, so that the content of the textbook is closer to real life, so that students have a sense of familiarity with the teaching content, and then have a thirst for knowledge, changing passive listening to active exploration. Therefore, if the teacher carefully analyzes the teaching content and designs teaching activities before class, and collects materials related to topics such as pictures, videos, or animations to show to students, it will greatly stimulate students' interest in learning. The design process of teaching activities is shown in Figure 2.

With the rapid development of modern technology, multimedia technology has been widely used in various disciplines. Grammar teaching reform should attach importance to creating simulated situations by multimedia means, activating the content of teaching materials, and combining language forms with language meanings, so that students can perceive and practice grammar knowledge in a pleasant environment and improve their ability to communicate in English (Lew et al., 2006). Multimedia technology is introduced into the teaching of English grammar in high school, and through research, its application in teaching is explored to seek a more effective way of teaching grammar that meets the needs of the times.

English Teaching Optimization Design Based on Content Video Retrieval Technology

Learning language itself is a boring and tedious thing, and students can only gradually realize the joy of learning on the basis of pleasure and good learning. It can be seen through the survey, students are very interested in the form of multimedia, which is determined by its unique intuitive, graphic, and vivid form of expression. It can not only display text, pictures, images, and sound in one, but also activate the content context (Datta et al., 2005). When you talk about it, you can also move static pictures like animations and present changing images like flowing water. So as to create a lively and

Figure 2. Teaching activity design process



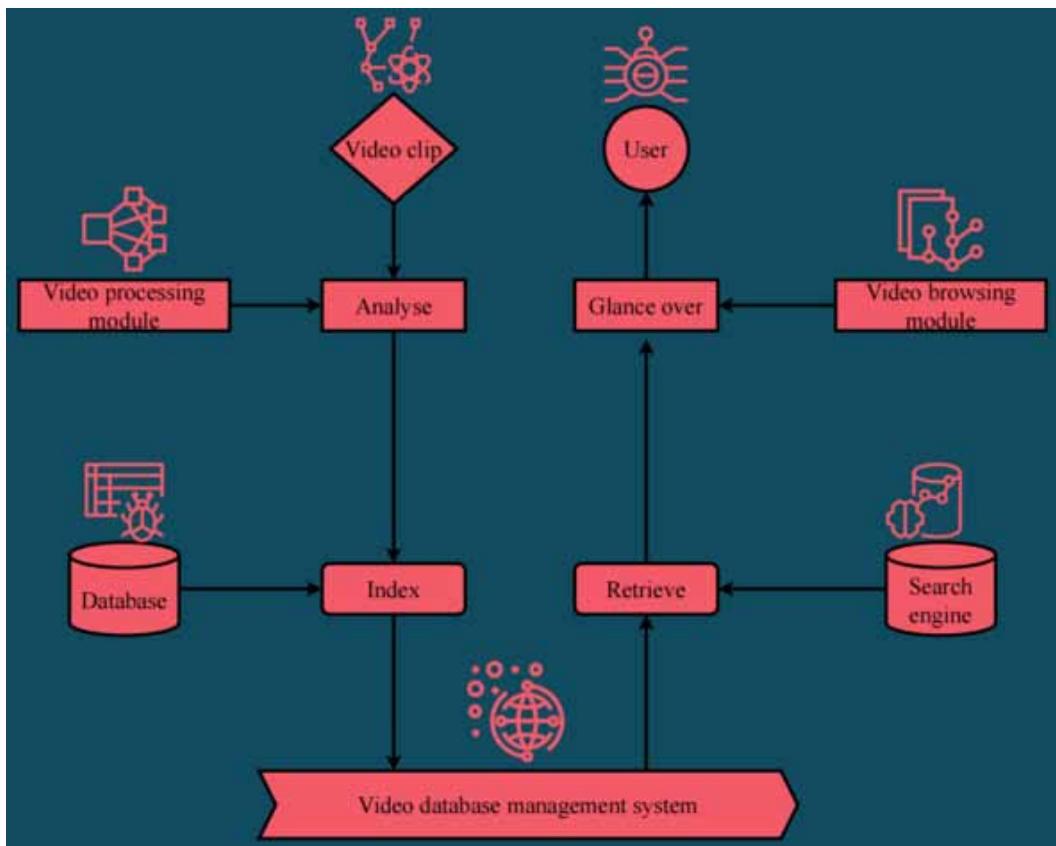
interesting language teaching situation for teaching, activate students' thinking, promote students' understanding of learning content, greatly reduce the difficulty of language learning for students, visualize abstract problems, simplify complex problems, and finally improve students' interest in learning. For example, the new edition of the third and fourth textbooks of college English not only has a long text, but also increases the number of words in each class, and thus the knowledge of grammar is relatively increased. At the same time, in order to better link up the knowledge, teachers have to review the related contents of the first two volumes, which have a lot of content and are difficult. In order to finish the teaching task on time and efficiently, multimedia courseware can be used to compare the previous content. Conventional teaching requires eight class hours, and multimedia courseware is used for classroom teaching. Video information is usually a diversified file recorded and edited by a camera or an image acquisition device and stored on a computer. The video file contains image information, audio information, and marked text information.

The characteristics of video information are continuous in time and moving pictures. From the perspective of the logical hierarchy of video, shots are video logical units on video frames. The video frame is the original data of video analysis. It is easy to obtain each frame of the video, but a single frame does not contain advanced semantic information. Therefore, in video analysis and processing, it is a key first step to divide the video into independent shot units. Video objects have their own

attributes and attribute values, which usually correspond to a set of spatially separable video frame sequences. This video retrieval system mainly includes two subsystems: video analysis and processing subsystem and content-based video retrieval subsystem. The detailed framework of the video retrieval system is shown in Figure 3.

Multimedia retrieval technology based on content is becoming more advanced and is gradually being applied to different retrieval databases. It can make it easier for consumers to easily access the vast multimedia information database and get the information they require. The use of content-based multimedia retrieval technology will expand due to the quick growth of network technology, which allows users to access multimedia information stored in databases all over the world. A novel retrieval technique, content-based multimedia retrieval technology, has only recently come into existence. The aims and scope of data retrieval are becoming more diversified and sophisticated as a result of the Internet's expansion. The retrieval technique must also find a solution to the issue of integrating with other retrieval technologies to increase retrieval efficiency if it is to become more commonly employed. The video processing subsystem mainly analyzes and processes the video files uploaded by users, and the specific analysis and processing mode is to analyze and process the video according to the analysis strategy put forward in the previous section. Its main function is to extract the theme subtitles, key frames, and divide the video into independent story segments and store this information in the database. Then, an index is established according to the obtained characteristic values. These are all implemented by the video processing subsystem. What users need is to upload video files to the system. The main function of a video retrieval subsystem is to provide users with retrieval interfaces

Figure 3. Frame diagram of video retrieval system



and browsing windows. The experimental platform mainly provides the retrieval methods of themes, key frames and video clips. The emphasis is to experiment the hierarchical feedback retrieval method based on theme, key frames, and video clips in the experimental platform. Video warehousing is the data entry way of video retrieval, and video retrieval is the retrieval function provided to users. In the implementation of the contents of these two functional units, different systems vary widely. But the general framework of video retrieval systems is basically the same.

The simplest and most direct method to detect abrupt shot transition is pixel pair comparison. This method involves comparing the gray difference of pixels in the corresponding positions of two adjacent frame images $f(x, y, t)$ and $f(x, y, t + 1)$, and when the number of pixels whose gray difference is greater than the set threshold exceeds a certain number, it is determined that there is a lens mutation between the two frame images. The formula for pixel comparison method is:

$$D = |f(x, y, t) - f(x, y, t + 1)| \quad (1)$$

Among them, x indicates the number of rows in the frame image, y indicates the number of columns in the frame image, t indicates the time axis position of the image frame, and D is the difference between two images. This method of pixel pair comparison is not suitable for camera and object motion. When the motion is large, the difference between two adjacent frames will exceed the set threshold, resulting in false detection.

Histogram comparison method is a statistical method. It is to calculate the histogram $H[f(x, y, t), k]$ of image frames, where $k = 0, 1, \dots, K - 1$, and compare the corresponding statistics of histograms of adjacent image frames to obtain their brightness difference. The calculation formula is:

$$D = \sum_{k=0}^{K-1} |H[f(x, y, t), k] - H[f(x, y, t + 1), k]| \quad (2)$$

If their brightness difference D is greater than the set threshold, it is determined that a shot abrupt change has occurred between the two frames. When determining the threshold, the difference between frames and the distribution of the difference should be calculated for the whole video. Let the mean value of the difference be m , and the standard deviation be σ , and the threshold value be T . Where $T = m + \sigma$ is usually set to a smaller number.

The x^2 detection method is to compare the distance between the components of the color histogram with the normalized x^2 experiment, D :

$$D = \sum_{k=0}^{K-1} \frac{\{H[f(x, y, t), k] - H[f(x, y, t + 1), k]\}^2}{H[f(x, y, t), k]} \quad (3)$$

This distance is used to judge the similarity. When the similarity is small and the threshold is set, it is determined that sudden changes have occurred between frames. This method is also affected by local illumination and target motion. Using the same processing method as above, the blocking method can avoid the influence of local illumination and target motion.

When N_{in} and cN_{out} have abrupt peaks, it indicates that shear has occurred at this time. When there is no abrupt peak, but there are continuous large values, it indicates that there may be a gradual

change. By comparing the relative values of N_{out} and N_{in} according to the detection criteria in Table 2, the fade-in, fade-out, and fold-over modes can be detected.

In this detection method, two thresholds are defined, one for detecting abrupt shot changes and the other for detecting gradual shot changes. Calculate the difference between the gray histograms of two adjacent image frames in turn. When the difference is greater than T_c , it is judged that abrupt shot transition has occurred between the two frames. And if the difference between the two frames is less than T_c , but greater than the threshold value of T_g , it is determined that the starting frame that may release the gradient is detected.

This double threshold detection method based on difference histogram is an effective shot recognition method. It can detect abrupt shot and gradual shot at the same time, avoiding the second shot detection processing of video. And this method can extend many accurate and efficient video shot recognition methods.

Content-Based Information Retrieval Method

Content-based video retrieval techniques and image retrieval techniques have many similarities. They both require feature extraction to transform the video or image into a computer-processable representation of digital features. Similarity computation is essential in order to compare the similarity between the query object and the media content in the database. To improve search efficiency, building an indexing structure is important for both. The user initiates a search request by entering query criteria and obtains results with high similarity to the query object.

Video retrieval techniques can utilize the visual features of video data to build an effective index. These features include key frames, shots, motion features, etc. Key frames can be used as a representation of video footage for quick and easy still image processing or video browsing. Motion-based retrieval methods, on the other hand, focus on utilizing the temporal characteristics of lenses and video objects in order to query camera motion information, scene motion, etc. By retrieving the direction and magnitude of motion of moving objects, relevant video clips can be obtained. Content-based video retrieval techniques and image retrieval techniques have similarities in feature extraction, similarity computation, index building, and query interfaces, etc., which can be borrowed and supplemented from each other, and jointly applied in practical scenarios.

According to the previous analysis, video can be organized into a set of different levels of units: frames, shot units, key frames, scenes, and story units. Then, according to the logical structure level of video, researchers put forward such a retrieval method: according to the video examples submitted by users, the video examples are structured into shot units, key frames, and scenes, and then the video examples are compared with the videos that have been processed as a result in the video database from these levels. This helps obtain the similarity between the video examples and the retrieved video data, and finally the videos with the similarity within the threshold range are returned to the users (Singhai & Shandilya, 2010). In the regular English classroom, when explaining texts and key grammar and sentence patterns, experienced teachers will leave enough time for students to think, and present the analysis of grammar or sentence patterns to students in the form of writing on the blackboard, while the transformation and unfolding process of other example sentences will be imagined and thought by students themselves. The blank space left by teachers on the blackboard is the space for students to think freely, and students can analyze the structure and characteristics of language according to the

Table 2. Detection criteria of fade-in/fade-out and folding

Gradient type	Fade in	Fade out	Superimposition
The relationship between N_{in} and N_{out}	$N_{in} > N_{out}$	$N_{in} < N_{out}$	$N_{in} > N_{out}, N_{in} < N_{out}$

teacher's performance on the blackboard (Meghini et al., 2001). At the same time, teachers cannot understand the students' mastery and cognition of information, and adjust the teaching progress and speed in time and place. Students cannot be encouraged synchronously, and they cannot experience the personality charm of excellent teachers. In the classroom teaching, the computer excessively displays the virtual situation, which makes the students have a distance from the real situation. The students blindly immerse themselves in the computer, alienating the emotional communication with the teachers, thus seriously affecting the quality of classroom teaching. Therefore, the relationship between multimedia and traditional media should be complementary. Multimedia teaching is an advanced and effective teaching method, but it does not mean that the traditional teaching mode can be completely abandoned. Only the reasonable integration of traditional teaching and multimedia teaching can better play their respective roles and advantages in teaching, so as to optimize teaching methods.

Now suppose there are two video files, in which video $iVideo$ contains $iFrame$ image frames, and video $jVideo$ has $jFrame$ image frames. The simplest way to detect whether videos $iVideo$ and $jVideo$ are similar is:

$$simijVideo = \sum_{i=1}^{iFrame} \sum_{j=1}^{jFrame} Dist(Hist(iVideo, i), Hist(jVideo, j)) \quad (4)$$

Among them, $Hist(iVideo, i)$ represents the histogram of the $iVideo$ frame image of video i , $Hist(jVideo, j)$ represents the histogram of the $jVideo$ frame image of video j , and $Dist()$ represents the difference between the two image histograms. If $simijVideo$ is larger than the similarity threshold set by the system, the two videos are considered similar, and if it is smaller than the similarity threshold, they are not. It can be seen from equation (4) that if you directly use equation (4) to calculate the similarity between two videos, the amount of calculation required is large, and the calculation time of the algorithm increases with the increase of the video file. It is acceptable to only compare the similarity between two videos, but it cannot be accepted when computing the similarity with massive videos in the database in turn. The system using this algorithm must also be unavailable.

The similarity comparison of any two video shots is converted into the similarity comparison between their corresponding 35-dimensional centroid feature vectors \bar{K}_1 and \bar{K}_2 . For the 32-dimensional color feature components in the 35-dimensional centroid feature vector, since they are normalized to between 0 and 1 and defined in the same physical domain, the Euler distance can be used to calculate the similarity of color features.

$$Similarity_c = \sum_{i=0}^{31} Min(\bar{K}_{1colorhistvgram}(i), \bar{K}_{2colorhistvgram}(i)) \quad (5)$$

For the other three texture feature components, each value is in a different range. Therefore, each numerical value is first normalized to the same range with Gauss during calculation.

$$Texture_i = \frac{Texture_i - m}{3\sigma} \quad (6)$$

$$Similarity_T = 1 - \frac{Similarity_T + 1}{2} \quad (7)$$

$$Similarity(\bar{K}_1, \bar{K}_2) = W_c * Similarity_c + W_t * Similarity_T \quad (8)$$

The Euler distance is used to calculate the texture similarity, and equation (7) is used to convert it to [0,1]. Finally, according to the weight W_c and W_t of color and texture components, use equation (8) to calculate the similarity between shots.

Assuming that the similarity between a shot V_q in the video $S_{q_i}(i = 1 \sim n)$ and all the shots in the retrieved video database (assuming that there are K shots, these K shots are composed of all the shots contained in the retrieved video data) is, the following thresholds are defined:

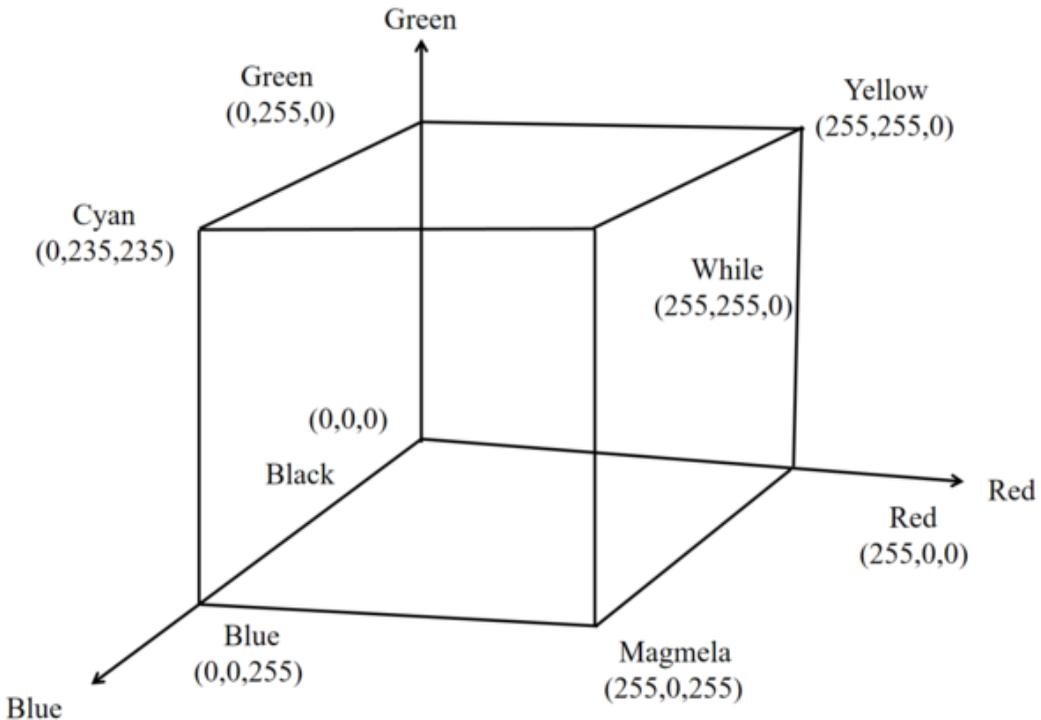
$$Threshold = \left\{ \begin{array}{l} \text{Similarity} \\ i \end{array} \left| \text{Max} \left(\frac{\sum_{j=i+1}^K \text{Similarity}}{K-i} - \frac{\sum_{j=1}^i \text{Similarity}}{i} \right), i = 1, 2, \dots, K \right. \right\} \quad (9)$$

Then, according to the threshold calculated by formula (9), if the similarity value between a shot V_q in the video $S_{q_i}(i = 1 \sim n)$ and the shot in the retrieved video database is greater than the threshold, it is determined that these shots are similar, otherwise, it is considered similar.

RGB color space is the most common color space used in computer image processing. Any color in the nature can be composed of R, G and B. RGB is a three-dimensional space. Different combinations of R, G, and B values result in different colors, that is, waves of different wavelengths. Each pixel in RGB image can be represented by the first quadrant point of 3D space, as shown in Figure 4.

Human skin color is in a small range in RGB color space, and the color display devices that are finally collected by existing image acquisition devices use RGB values. However, RGB color space

Figure 4. RGB color space



is not intuitive and perception is very uneven. Therefore, it is difficult to establish an accurate skin color model in RGB color space. The general processing method is to convert RGB image to a certain color space, such as HSV, YCbCr.

If the similarity of shot levels is measured by the three characteristic values of feature f_{sk} , color, texture and time span, then $k = 3$, that is, *Sity C, SimilarityT, time_spanning*, corresponds to three weights of $W_{st} = (W_c, W_t, W_{t_s})$ respectively.

The concrete calculation method is to add up the Set_i values of S_j^i in *SimilarityC, SimilarityT, time_spanning*, and then find their average values, where each one corresponds to a ternary group, where

$$average_Similarity_c^i = \sum Similarity_c / N_i \tag{10}$$

$$average_Similarity_t^i = \sum Similarity_t / N_i \tag{11}$$

$$average_time_spanning_c^i = \sum time_spanning / N_i \tag{12}$$

Then m , Set_i forms a matrix of $m \times k$, $k = 3$. Then, the variance of each column is calculated. The variance represents the similarity of the components in this column. If the values of a certain component are more similar, the greater their contribution to the overall similarity, that is, users think that this feature is more important to the lens similarity.

So you can get: $W_{sk} = 1 / \sigma_k$. Where W_{sk} corresponds to W_c, W_t, W_{t_s} respectively. The final calculation of W_{sk} is:

$$W_{sk} = \frac{1}{\sum W_{sk}} \tag{13}$$

At this point, the weight values can be updated to measure the shot-level similarity. The weights of the color and texture components account for a small percentage of the video similarity, so there is no need to consider the weights of the color and texture components.

Retrieval is the processing of comprehensive query, description, matching, and extraction. The query may or may not succeed. According to the principle of pattern recognition, four cases can be obtained as shown in Table 3.

Assuming that N_c indicates the correct number of videos in the retrieval result set, N_f indicates the wrong number of videos, and N_m indicates the number of videos missed during retrieval, the recall and precision are defined as follows:

$$recall = \frac{N_c}{N_c + N_m} \tag{14}$$

Table 3. Basic Parameters of Retrieval System Performance

	Related	No association
Retrieved	A (correct retrieval)	B (error retrieval)
Not retrieved	C (missing inspection)	D (correct rejection)

$$precision = \frac{N_c}{N_c + N_f} \quad (15)$$

Where, $I_j^*(x, y)$ indicates the pixel brightness at the position i in the subtitle area block I_j of the (x, y) frame. Then quantize to a value of 0,1 with the following formula:

$$r_i(x, y) = \begin{cases} 1 & \text{if } \Delta I_j^*(x, y) \geq \beta_{D1} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Where β_{D1} is the set pixel brightness change threshold. Let $\mu_i(j)$ indicate the pixel brightness change of a small square at the corresponding position of two adjacent frames, and then use the following formula to calculate:

$$\mu_i(j) = \sum_{(x,y) \in B_j} r_i(x, y) \quad (17)$$

The subtitle information of video frames also includes scrolling news subtitle information and report subtitle information. In extracting the subject subtitles, the scrolling news subtitles and report subtitles can be filtered out based on whether the subtitles have a relative offset between neighboring frames and the duration of the subtitles in the video.

RESULT ANALYSIS AND DISCUSSION

For the perception of English teaching information, in the non-Internet era, many scholars have conducted research on the perception of teaching quality. At the same time, some scholars have classified the cognitive perception of English teaching into the category of quality perception. However, in the Internet era, many teachers are vulnerable to the influence of external information. The reason for this phenomenon is that the rapid spread of external information has changed teachers' cognition of English teaching. Although the influence path of cognitive perception is similar to that of quality perception, they are essentially different. This paper analyzes the impact mechanism and evolution process of these two types of information perception. This study has strong theoretical value in the impact of information perception on consumer demand. At the same time, in the Internet era, the relevant conclusions of this study can be used in the strategies of universities to achieve their own teaching goals. Therefore, this paper has a strong practical significance.

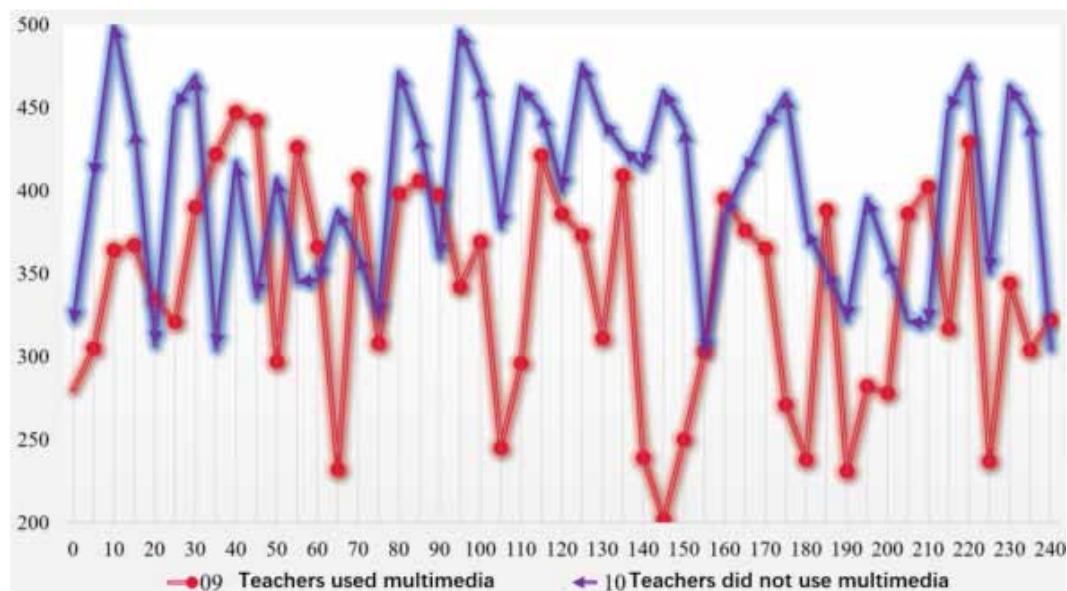
After the multimedia courseware was made, two classes of students majoring in international trade in higher vocational colleges of our college were selected as the objects of the application experiment of the multimedia courseware of New Vocational English. The two classes adopted different teaching modes and achieved different teaching results. For example, the warm-up part before class: in this part of the experimental class, the traditional warm-up method has been changed. Instead of explaining the subject content of the teaching unit in language, a dynamic picture is presented with multimedia courseware. Four different dynamic interfaces are designed and displayed, with the unit as the theme. In each key link, students can conduct simulation exercises according to their actual situation. After the exercise, students can play roles in groups in class. Multimedia courseware can help students intuitively understand English language knowledge and make full use of multimedia technology to create a good language communication environment and make students feel the stimulation of language in a real language environment.

In the process of e-teaching, it is necessary to make full use of modern multimedia technology and explore new teaching modes, so as to improve the effectiveness of e-teaching in higher vocational colleges and universities. After the completion of the multimedia courseware, I selected two classes of students majoring in international trade in the higher vocational colleges of our college as the objects of the application experiment of the multimedia courseware of New Vocational English. The two classes used different teaching modes to teach and achieved different teaching effects. For example, the warm-up part before class: in this part of the experimental class, the traditional warm-up method was changed. Instead of explaining the theme content of the teaching unit through language, multimedia courseware was used to present a dynamic picture. Four different dynamic interfaces are designed and displayed with the theme of the unit. In each key link, students can carry out simulation exercises according to their own actual situation. After the exercises, students can use role play in the form of group to show in class. Multimedia courseware can help students understand English language knowledge visually.

Based on the experimental data of this study, statistics and calculations were conducted. Figure 5 is a comparative analysis of teaching effectiveness, with the horizontal axis representing different teaching modes and the vertical axis representing students' academic performance and satisfaction. The data comes from the final exam scores and questionnaire survey results of the experimental and control classes. From the figure, it can be seen that the experimental class using multimedia courseware outperforms the control class using PPT courseware in terms of academic performance and satisfaction. As shown in Figure 5 below, it can be seen that students in the classroom interact with each other, actively participate in discussions and speeches, take a quiz on classroom sentence patterns and investigate extracurricular activities. The participation of students in the experimental class using multimedia courseware is higher than that in the comparison class. This shows that multimedia teaching mode can make students have a strong interest in learning, thus mobilizing students' inner needs and providing students with a more free and relaxed space.

The students thought that the teaching content of the course was closely combined with the future practical work, and the classroom teaching atmosphere was good, which could arouse

Figure 5. Comparative analysis of teaching effect



everyone's enthusiasm for learning. Moreover, it enhances students' ability to analyze and solve problems, so that they have confidence and grasp of their professional ability in future job applications.

After the screen theme subtitles are detected, OCR is used to identify the news theme subtitles after binary processing, so that the news theme subtitles can be accurately identified. Using the theme detection and recognition method proposed in this paper, a group of "military observation room" videos are analyzed experimentally. The experimental results are shown in Figure 6. Through experiments, it can be verified that the average accuracy of this method can reach 92.45%.

Figure 6 shows the accuracy analysis of video topic detection. The horizontal axis represents different video categories, and the vertical axis represents the accuracy of topic detection. There are mute clips in the teaching video, and these mute clips generally appear between two adjacent news story units. Therefore, when clustering shots into news story units, the time when the mute frequency band appears can be used as the dividing line between adjacent story units. Therefore, it is very important to find out the silent clips in teaching videos for the division of news story clips.

From the analysis of Figure 7, students' recognition of the teaching content of multimedia courseware that highlights professional characteristics is strong in practicality, and highlights key and difficult points is higher than that of PPT courseware. The evaluation of interaction, animation design, and interface design of courseware on the technical level is obviously higher than that of PPT courseware. However, the evaluation of courseware structure design and ease of operation is lower than that of PPT courseware. It is concluded from the figure that the multimedia courseware needs to be improved in terms of facilitating students' operation and structural design.

Figure 7 is an analysis of multimedia courseware operations, with the horizontal axis representing different courseware content, and the vertical axis representing students' recognition of the courseware content. The data is sourced from the questionnaire survey results of the experimental class students. From the figure, it can be seen that students have a higher recognition of multimedia courseware teaching content that highlights professional characteristics, strong practicality, and highlights key and difficult points than PPT courseware. Based on the above analysis, multimedia assisted English Language Teaching has obvious advantages in improving learners' learning

Figure 6. Analysis of video topic detection accuracy

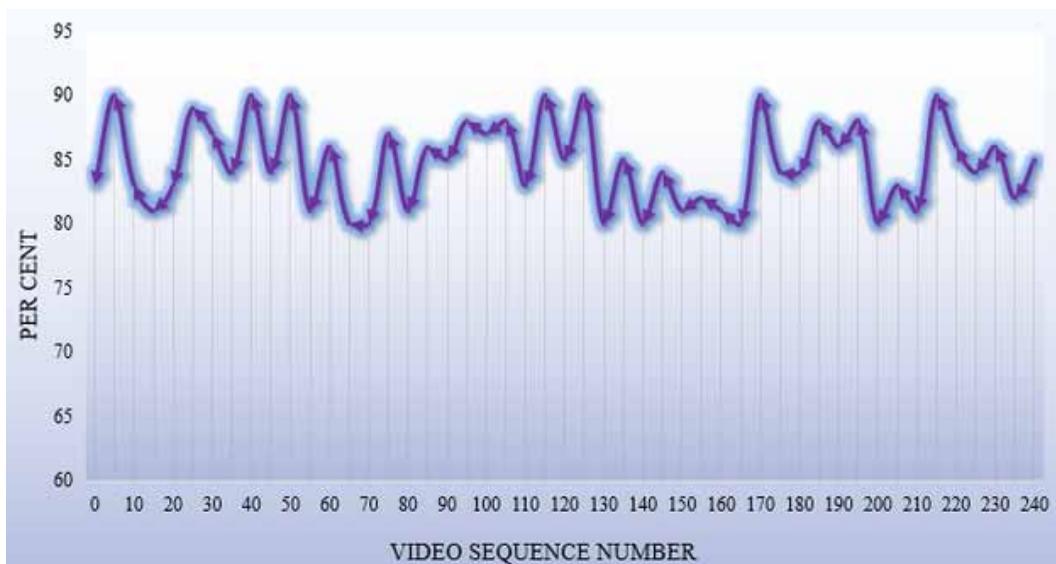
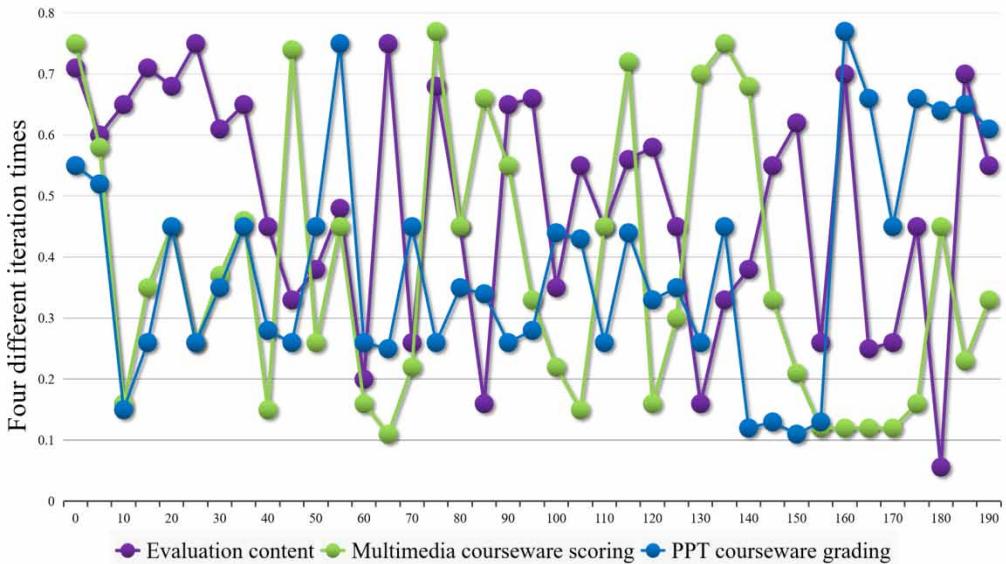


Figure 7. Operation analysis of multimedia courseware



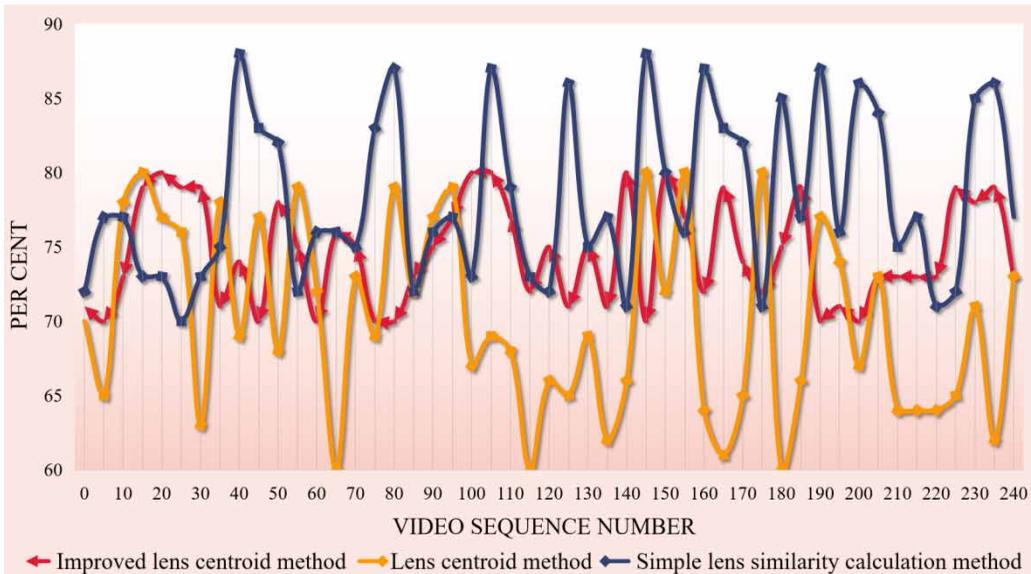
behaviors and attitudes. It is embodied in the following aspects: First, students’ learning attitude has changed. The multimedia courseware of New Professional English is novel, lively, and rich in content. It has designed various forms of exercises related to the future workplace, both visual and abstract, which can meet different learning requirements. Second, the ability of autonomous learning has been improved. The role of multimedia is exaggerated, which makes some teachers pay too much attention to multimedia. For the classroom, the realization of multimedia teaching is based on the success of courseware, so the teaching quality is an important factor to promote the progress of multimedia teaching.

In this paper, the lens center point algorithm, the improved lens center point algorithm, and the simple video clip similarity algorithm are used to calculate the similarity of 10 groups of video samples, respectively. The horizontal axis represents the video group number, and the vertical axis represents the similarity of videos between groups. The experimental results are shown in Figure 8. The experiment shows that the improved shot centroid algorithm is superior to the other two methods in the accuracy of similarity calculation. Using a simple video segment similarity algorithm, it is easy to see the difference between two frames. It is very sensitive to motion and easy to cause false detection. The shot centroid algorithm is not very sensitive to motion, although it loses the correlation between images. The adjacent frames in the same shot generally have the same elements of global vision. That is, adjacent frames in the same shot have similar color distribution. It is reflected in the figure that the graphic difference between adjacent frames in the same shot is small. The graphics between frames in different shots are quite different.

Based on the comparison results in Figure 8, the differences between the three video similarity calculation methods can be seen. The horizontal axis represents the video group number, and the vertical axis represents the similarity of videos between groups. The data comes from 10 video samples for which similarity calculation was performed.

From the figure, it can be observed that the improved lens centroid algorithm shows a better advantage in the accuracy of similarity computation over the other two methods. This indicates that the method proposed in this paper is able to evaluate the degree of similarity between videos more accurately.

Figure 8. Comparison results of three video similarity calculation methods



Based on the results in Fig. 8, it can be concluded that the improved lens center point algorithm outperforms the other two methods in terms of accuracy in video similarity calculation. Teaching students in accordance with their aptitude is a universal law of teaching, and its essence is to achieve targeted teaching centered on students. In college ET, due to the individual differences in the basis and level of students' English knowledge, the level and ability of individuals to receive information, and digest knowledge, different students will have different choices for language carriers and practice forms in the process of foreign language learning. This requires teachers to organically organize and integrate the teaching materials with different levels and requirements, such as texts, exercises, questions, answers and tests, when designing multimedia courseware and organically organize the teaching requirements and different teaching materials related to these teaching contents, which can not only enable students at different levels to gain in different degrees, but can also aim at their different defects in listening and speaking, so as to further promote the optimization of ET. Under the guidance of the theory of teaching optimization, the new classroom teaching mode of multimedia is used to better realize the optimization of college English Language Teaching through the process of learner centered teaching, such as situation creation, interactive teaching, and individualized teaching.

To sum up, in English classroom teaching, the reasonable use of computer-assisted multimedia teaching system can solve the audiovisual and other related problems in English teaching. Students can develop their audio-visual ability through audio-visual perception of English language materials. At the same time, teachers play a leading role in the classroom to guide students to master the basic knowledge of English and train the basic skills of the language. In this way, students can cultivate their ability to think and express themselves in English and lay a solid foundation for realizing communication in English.

Through the introduction of computerized multimedia-assisted teaching systems, English teaching methods have been enriched and the quality of teaching has been improved. The increased demand for foreign language teaching also promotes the development of foreign language teaching media. The future of education is dominated by high technology, and computer multimedia intelligent education becomes characteristic. In the implementation of science and

education today, the development of education in China urgently needs the perfect combination of high-tech intervention and current teaching.

Therefore, the rational use of multimedia teaching in English teaching can greatly improve classroom efficiency. English teachers should strive to improve the ability to use modern educational technology, master the skills of courseware production, and understand the basic rules of multimedia use in teaching. Language education must involve scientific integration of teaching objects and teaching content and strive to achieve the best teaching effect.

CONCLUSION

Based on the experimental results and data analysis of this study, it can be concluded that the application of multimedia information retrieval technology in an English teaching management system is very effective and can significantly improve students' learning behaviors and attitudes, increase students' interest in learning and motivation to learn, which is of great significance in improving the quality of the course and cultivating students' abilities. In addition, compared to traditional PPT courseware, multimedia courseware is more easily accepted and liked by students. Meanwhile, this study also verifies the accuracy of the camera centroid algorithm, the improved camera centroid algorithm, and the simple video clip similarity algorithm, and puts forward relevant solutions and suggestions to further improve and optimize the multimedia-assisted English teaching system to enhance the teaching effect and students' performance. Starting from the monism relationship between information perception and quality perception, this paper studies and analyzes the relationship between the selection of publicity and teaching quality in colleges and universities. According to the game theory, the university teaching model is established. From the perspective of this model, the stability of universities and students is analyzed. It provides a theoretical basis for university behavioral choices. Under the influence of this process, the evolution mechanism of the two types of information in the transmission process is analyzed and studied, and the evolution process of market demand is analyzed and studied. On this basis, a cognitive perception communication influence model based on the maximization of influence is proposed to calculate how colleges and universities can achieve the maximum growth of English teaching demand with less investment. The system is a network teaching and evaluation system, which can be used as a subsystem of the campus management system. It is convenient for teachers to release course information, teaching courseware, online homework and correct students' homework, provide students with regular examinations, and complete teachers' and students' evaluation of teaching. The establishment of perception model in English teaching is mainly based on the perception model between students and teachers. Through the communication between students and teachers, and the integrated communication, it is conducive to the combination of classroom and practice. This leads students to deeper understanding and expanding their thinking space and encourages students to learn more independently and improve teaching efficiency.

The development of the perception model is still beset by issues including a lack of experts, ineffective operations, low intelligence, a lack of information exchange capacity, and excessive expenses. To further integrate the Internet of Things with educational practice, more thorough research is still needed on the application of the English perception model on campuses.

DATA AVAILABILITY

The figures and tables used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGEMENTS

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

- Albiladi, W. S., & Alshareef, K. K. (2019). The application of blended learning in ET and learning: A review of existing literature. *Language Teaching Research*, *10*(2), 23–28.
- Alshorman, B. A., & Bawaneh, A. K. (2018). Attitudes of faculty members and students towards the use of the learning management system in teaching and learning. *Turkish Online Journal of Educational Technology-TOJET*, *17*(3), 1–15.
- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. (2019). The use of ontology in retrieval: A study on textual, multilingual, and multimedia retrieval. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 21662–21686. doi:10.1109/ACCESS.2019.2897849
- Benavent, X., Garcia-Serrano, A., Granados, R., Benavent, J., & de Ves, E. (2013). Multimedia information retrieval based on late semantic fusion approaches: Experiments on a wikipedia image collection. *IEEE Transactions on Multimedia*, *15*(8), 2009–2021. doi:10.1109/TMM.2013.2267726
- Buckner, E., & Kim, P. (2014). Integrating technology and pedagogy for inquiry-based learning: *The Stanford Mobile Inquiry-based Learning Environment (SMILE)*. *Prospects*, *44*(1), 99–118. doi:10.1007/s11125-013-9269-7
- Burgin, X., & Daniel, M. C. (2017). Exploring English language teaching in an Ecuadorian urban secondary institution. *GIST–Education and Learning Research Journal*, *(14)*, 107–134. doi:10.26817/16925777.364
- Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Journal of Educational Technology & Society*, *13*(2), 13–24.
- Dahl, V., Cukierman, D., Bel-Enguix, G., & Jiménez-Lopez, M. D. (2010, May). Logic programming: Teaching strategies for students with no programming background. In *Proceedings of the 15th Western Canadian Conference on Computing Education* (pp. 1-5). ACM. doi:10.1145/1806512.1806515
- Datta, R., Li, J., & Wang, J. Z. (2005, November). Content-based image retrieval: Approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* (pp. 253-262). ACM. doi:10.1145/1101826.1101866
- Du, W., & Liang, X. (2021). The application of multimedia courseware in teaching college English intensive reading from the perspective of constructivism. *International Journal of Electrical Engineering Education*, 0020720920984312.
- Habibi, A., Mukminin, A., Riyanto, Y., Prasajo, L. D., Sulistiyo, U., Sofwan, M., & Saudagar, F. (2018). Building an online community: Student teachers' perceptions on the advantages of using social networking services in a teacher education program. *Turkish Online Journal of Distance Education*, *19*(1), 46–61. doi:10.17718/tojde.382663
- Jitpaisarnwattana, N. (2018). Fostering learner autonomy in foreign language classroom: A digital storytelling project. *Journal of Foreign Language Education and Technology*, *3*(2), 136–160.
- Lauder, A. (2008). The status and function of English in Indonesia: A review of key factors. *Makara Human Behavior Studies in Asia*, *12*(1), 9–20. doi:10.7454/mssh.v12i1.128
- Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. [TOMM]. *ACM Transactions on Multimedia Computing Communications and Applications*, *2*(1), 1–19. doi:10.1145/1126004.1126005
- Liang, L. (2022). Teaching English as a foreign language using the Internet and multimedia environment. *Soft Computing*, *26*(20), 10891–10902. doi:10.1007/s00500-022-07144-2
- Liu, M., Navarrete, C. C., & Wivagg, J. (2014). Potentials of mobile technology for K-12 education: An investigation of iPod touch use for English language learners in the United States. *Journal of Educational Technology & Society*, *17*(2), 115–126.
- Meghini, C., Sebastiani, F., & Straccia, U. (2001). A model of multimedia information retrieval. [JACM]. *Journal of the Association for Computing Machinery*, *48*(5), 909–970. doi:10.1145/502102.502103

Mulyono, H. (2016). Using Quipper as an online platform for teaching and learning English as a foreign language. *Teaching English with Technology, 16*(1), 59–70.

Ottenbreit-Leftwich, A. T., Brush, T. A., Strycker, J., Gronseth, S., Roman, T., Abaci, S., vanLeusen, P., Shin, S., Easterling, W., & Plucker, J. (2012). Preparation versus practice: How do teacher education programs and practicing teachers align in their use of technology to support teaching and learning? *Computers & Education, 59*(2), 399–411. doi:10.1016/j.compedu.2012.01.014

Schmid, E. C. (2010). Developing competencies for using the interactive whiteboard to implement communicative language teaching in the English as a Foreign Language classroom. *Technology, Pedagogy and Education, 19*(2), 159–172. doi:10.1080/1475939X.2010.491218

Singhai, N., & Shandilya, S. K. (2010). A survey on: Content based image retrieval systems. *International Journal of Computer Applications, 4*(2), 22–26. doi:10.5120/802-1139

Syakur, A. (2020). The effectiveness of English learning media through Google Classroom in Higher Education. *Britain International of Linguistics Arts and Education (BIO LAE). Journal, 2*(1), 475–483.

Tsai, Y. R. (2015). Applying the Technology Acceptance Model (TAM) to explore the effects of a Course Management System (CMS)-Assisted EFL writing instruction. *CALICO Journal, 32*(1), 153–171. doi:10.1558/calico.v32i1.25961

Xu, H., & Yao, Y. (2022). Research on multimedia technology-assisted college English grammar teaching. *Wireless Communications and Mobile Computing, 2022, 2022*. doi:10.1155/2022/7534153

Jingfang Sun was born in Shandong, China, in 1979. From 1998 to 2002, she studied in Qingdao University and received her bachelor's degree in 2002. From 2005 to 2009, she studied in Ocean University of China and received her Master's degree in 2009. Currently, she works in Shandong Foreign Trade Vocational College. She has published ten papers and one book. Her research interests are English teaching and Cross-culture communication.