Computers in Human Behavior 41 (2014) 365–373

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Hyperaudio learning for non-linear auditory knowledge acquisition

Joerg Zumbach [a,*], Neil Schwartz [b]

[a] University of Salzburg, Austria
[b] California State University, Chico, CA, USA

ARTICLE INFO

Article history:

Keywords:
Hypermedia
Mobile learning
Hyperaudio

ABSTRACT

In this research, we present the concept of Hyperaudio as non-linear presentation of auditory information in the context of underlying theoretical assumptions of how Hyperaudio differs from existing non-linear information media. We present a study comparing text and auditory represented information either in a linear or non-linear manner and the interaction of these presentation formats with different underlying text types. Learners had to learn from two different text sorts either from text only in linear or non-linear manner from a computer screen or the same information presented as audio files also presented either in linear or non-linear manner. Results show overall advantages of linear information presentation compared with non-linear information presentation, and the advantages of written text versus auditory text on learning performance assessed with an essay task and a multiple-choice test. Interaction effects indicate that non-linearity increases cognitive load assessed with a self-report measure in auditory instruction compared to linear information presentation while cognitive load in processing written text is not affected by linearity. Further, effects reveal that the text type (ex-pository vs. linear text type) interacts with presentation format showing that expository text leads to comparable learning outcomes in linear and non-linear formats, while presenting linear text type as hypertext or Hyperaudio is here rather unbeneficial.

1. Introduction

In this study, we examined the influence of linear vs. non-linear Hyperaudio on learning outcomes, measured by two different metrics of knowledge acquisition, under the lens of cognitive load theory. The basic research questions were aimed at understanding how auditory and text-based learning material interact with information access (linear vs. non-linear) and different types of text.

The use of auditory instruction has increased during recent years (O'Bannon, Lubke, Beard & Britt, 2011; Vajoczki, Watt, Marquis, & Holshausen, 2010). A major reason for this increase is the increasing ubiquity of mobile audio devices like MP3-players or cell and smart phones. A common use of these mobile devices for learning purposes is to listen to instructional material designed exclusively for auditory instruction, like audiobooks or podcasts. This material is mainly characterized by a linear sequence of orally presented information.

In this paper, we suggest a different format of auditory instruction—specifically, nonlinearly-presented audio information—termed "Hyperaudio". Hyperaudio is comparable to non-linear visual learning material like Hypertext or Hypermedia and audio-visual non-linear learning environments like Hypervideo (e.g., Zahn, Schwan, & Barquero, 2002). The difference between linear text and Hypertext is characterized by the way information is presented and accessed. While linear text (usually presented on a digital display) can only be accessed in a linear manner with "next" and "back" buttons, Hypertext nearly always allows non-linear information retrieval by using interactive hyperlinks that connect different information resources in an associative manner (cf. Chen & Rada, 1996; e.g., like the hyperlinks available in Wikipedia articles). Interestingly, there is a large body of research related to Hypermedia learning and some about Hypervideo, but a dearth of research about Hyperaudio.

We define Hyperaudio as an arrangement of auditorily presented material represented within locally coherent hyperlinked nodes. These audionodes are connected via hyperlinks that enable users to navigate within a Hyperaudio environment. By navigating this Hyperaudio environment, users should be able to understand the relationships between single audionodes as well as develop understanding across nodes for a sense of global coherence. The possibilities of linking nodes and creating an overall navigation structure of a Hyperaudio environment are the same as in common hypertexts or hypermedia environments. Nodes can be linked in a

* Corresponding author at: School of Education, University of Salzburg, Hellbrunner Str. 34, A-5020 Salzburg, Austria. Tel.: +43 (0)662 8044 5801.
E-mail address: joerg.zumbach@sbg.ac.at (J. Zumbach).

http://dx.doi.org/10.1016/j.chb.2014.10.024
0747-5632/

linear, hierarchical, elaborative, or an associate manner (cf. Grabinger & Dunlap, 1996).

However, contrary to linear audiobooks, audio nodes are not visited in a predefined sequence. For example, when a learner listens on a cell phone to an audio document explaining bacteria, the document might be linked by the authors to several other audio files, like cell cores, viruses, etc. These links are displayed on the cell phone while the learner is listening to the document. If the learner is interested in any of the documents, the learner might simply activate the link (e.g., by pressing a corresponding button on the phone) and a new audio file would open. The new file would be linked to other audio files, and so on. In short, current standards and developments in Software (data reduction, e.g., MP3) and portable hardware (e.g., portable MP3-Players, or integrated audio-players in cellular phones) open a wide range of applications and can contribute to research on ubiquitous learning with handheld computer devices (e.g., Hsi, 2003; Roschelle & Pea, 2002).

With regard to hypertext, research on learning has shown that hypertext learning is not always beneficial. Although a meta-analysis provided by Chen and Rada (1996) revealed comparable outcomes or slight advantages of hypertext compared with linear text on learning, Shapiro and Niederhauser stated that "(...) a number of studies have not shown such an effect (...)" (2004, p. 609). Foltz (1996, see also Wells & McCrory, 2011) also argued that learning with non-linear texts increases complexity and makes it harder to learn. Factors like navigation planning, the structure of hypertexts, and learner prior knowledge also influence information retrieval (cf. Salmerón, Baccino, Cañas, Madrid, & Fajardo, 2009). Finally, Zumbach and Moharz (2008) added that while learner characteristics influence performance in hypertext learning compared with linear text, issues of instructional design are also important. They found that basic text design, text type, and complexity of learning material play an important role when comparing learning from text and hypertext. In particular, learners must compensate for poor readability of text when instructional design is poor. This puts a heavier load on working memory resources, which are limited in hypertext because of the dual task of information retrieval and navigation planning. In addition, some text types are traditionally designed in a linear sequence and follow a linear (time) plot. Breaking this linear plot in hypertext can also increase working memory load because learners have to re-construct the plot, which is not necessary in traditional text presentation. Thus, Zumbach and Moharz (2008) argue that complexity of content plays an important role in hypertext learning. When the complexity of a learning domain is high, learners have to make use of multiple information sources. Therefore, designers of a hypertext system might offer links between different aspects of a topic, which is impossible to do with traditional linear text. Taken together, research reveals an ambiguous picture showing that non-linear information media may foster, but also harm, learning depending on learner characteristics, the nature of content, and issues of instructional design.

2. Auditory vs. text-based instruction

Early comparative studies seemed to show an advantage for auditory instruction over text-based instruction (cf. Golas, Orr & Yao, 1994; Nugent, 1982). However, there were many constraints of these early investigations because of the role and behavior of the reader, in addition to shortcomings in the methods of data analysis. More recent research presents ambiguous outcomes (for an overview see Barron, 2004). In fact, Travers (1970, quoted after Barron, 2004, p. 954) stated that "One cannot reasonably ask the general question whether the eye or the ear is more efficient for the transmission of information, since clearly some information is better transmitted by one sensory channel than by another". The processing of either text or audio is also strongly influenced by the proprietary design of each modality. That is, within text, several design features like headings, paragraphs, highlighting, etc. can support reading comprehension processes that are not easily transferable to spoken language (cf. Hartley, 2004). In addition, reading and rehearsal strategies are made easier by self-pacing during the encoding of information presented in text, but it is difficult to engage in the same processes when information is presented auditorially. In spoken language, there are also stylistic features that can contribute support to comprehension processes. These features consist of the control of speed, use of breaks, emphasis on certain information, etc. that are not available in written text (cf. Barron, 2004; Kürschner & Schnotz, 2008). Another advantage of auditory instruction is that spoken language contains additional information to the message itself. For example, voice carries paralinguistic personality cues that are referred to the speaker (e.g., Nass & Lee, 2001; Nass & Brave, 2005). Common to both modalities is the representation learners create from the text base; that is, learners construct mental representations based on the propositional text model, and this representation is modality *unspecific* (cf. van Dijk & Kintsch, 1983; Johnson-Laird, 1983).

3. Working memory and processing of text and audio

The complexity between visual and verbal material can be informed by an analysis of the way learners process information in working memory (WM). According to Baddeley's (1992, 1998) WM model, auditory information is processed in a phonological loop where it is stored and shortly repeated before subsequently being processed within an episodic buffer. Visual text-based information is first represented in a visuo-spatial sketchpad before being processed within the episodic buffer. Rummer, Schweppe, Scheiter, and Gerjets (2008) assume that text and audio are likely to be represented only in the phonological loop, arguing that Baddeley's model is not modality-specific but rather codality-specific. That is, verbal information processing takes place in the phonological loop independent from its modality; visual-verbal information is imported in the phonological loop by rehearsal. In short, WM load should be independent of modality. If Rummer et al. (2008) are correct, then differences between reading and listening comprehension should not be expected.

And yet, several studies comparing learning with written or auditory verbal information have shown differences between both modalities. For example, there seems to be slight advantages for auditory information compared to visual information, because visual information more often contains irrelevant and distracting stimuli that may have less semantic meaning but still has to be processed (e.g., word or number lists; cf. Kürschner & Schnotz, 2008; Pächter, 1996). For text comprehension, there seems to be advantages of written learning material regarding memory for details, while auditory presentations seem to contribute to better understanding of more overarching concepts or ideas—like a plot, for example (e.g., Hildyard & Olson, 1978; Kürschner, Schnotz, & Eid, 2006).

It is possible that a major advantage of complex text comprehension is due to self-pacing and the application of specific reading strategies—both of which might explain advantages of written text compared to auditory text—at least with regard to the memorization of details. Indeed, working memory capacity is stressed with auditory presentation formats, especially when texts are longer or more complex and contain information that is redundant. Leahy and Sweller (2011) underscore this difference in advantage in their discussion of the transient information effect—an effect characterized by a learner's tendency to drop information kept in

WM in favor of newly gained information when the aggregate of both information sources exceeds working memory capacity. Thus, it is easier for readers, compared to listeners, to go back to a certain passage within a text and refresh that information. Auditory information, on the other hand, may stay rather transient, because it is much more complicated to navigate within an audio file. Taken together, an exclusively visual presentation, in contrast to an auditory one, might contribute to decreased working memory load, because information retrieval is simplified, leading to better overall comprehension (Tindall-Ford, Chandler, & Sweller, 1997).

4. Research questions and hypotheses

Based on the discussion above, research is mixed on the relative value of visually- and auditorially-presented text. Auditory learning seems to primarily support global coherence construction and text-based learning seems to support local coherence formation. However, to date, there is no research on how this difference in coherence construction might interact with the non-linearity of auditorially-presented information. Thus, the research questions were as follows:

First, research contributes evidence that written text supports learners in remembering detailed facts, while auditory learning material contributes to understanding higher order ideas (e.g., Kürschner & Schnotz, 2008) and more global coherence formation (e.g., Kürschner et al., 2006). Thus, we expected learners in the auditory condition to show better learning performance in measures of overall text comprehension—especially when the material was presented in a linear fashion (Hypothesis 1). Alternately, we expected learners in the text-based condition to show better performance on text-based details when the material was presented nonlinearly; learners in the auditory condition were expected to perform better when the text was linear.

We were also concerned with whether different representations (text vs. audio, linear vs. non-linear) would be affected by the nature of the learning material. Theoretically speaking, we wondered how different text structures, which focus on local vs. global coherence formation, might influence learning outcomes. Recent research by Zumbach and Mohraz (2008) indicated that there is a strong interaction between the type of text and linear vs. non-linear information retrieval when learning from hypermedia or linear text. Learning in traditional text-based learning environments demands that learners first build up local coherence based on the information presented in a node of the hypermedia network and subsequently build up global coherence by integrating all the accessed information into a mental model (cf. Schnotz & Zink, 1997). Following these considerations, learners learning from a text-based linear learning environment should benefit in remembering single facts (because of local coherence formation); learners with linear auditory information should benefit in building global coherence, better remembering the main structure and ideas of the presented information. Thus, auditory information retrieval should lead to decreased learning performance compared to text-based non-linear learning environments when learning in a domain with a focus on single and unrelated concepts (Hypothesis 2). This latter effect is likely to vanish with a text that follows a strong plot.

Finally, differences between linear and non-linear learning environments have to be expected since non-linear navigation planning might increase extraneous cognitive load (ECL), based on the assumptions of Cognitive Load Theory (CLT; Paas, Renkl, & Sweller, 2004; Sweller, 1994). CLT differentiates between *intrinsic cognitive load* (ICL), *extraneous cognitive load* (ECL) and *germane cognitive load* (GCL; Sweller, van Merriënboer, & Paas, 1998). ICL results from direct information processing and is, thus, unavoidable. ECL is determined by the design of the learning materials. GCL is necessary to activate schemata in order to process new information more deeply.

Thus, ECL might be a part of a phenomenon called "cognitive overhead" in non-linear information retrieval (cf. Conklin, 1997; Zumbach, 2006). As such, navigation planning might not only be a compromising factor influencing knowledge acquisition, but it might also be necessary to support global coherence formation. Additional mental effort can also enhance learning if additional effort leads to deeper elaboration processes and coherence building.

A problem, however, is that it is impossible to identify directly whether higher mental effort due to increased cognitive load is advantageous or disadvantageous for learning simply by assessing cognitive load with self-report measures (cf. Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Therefore, we reasoned that by observing both learning performance and self-reported mental effort, it might be possible to distinguish between positive and negative influences of cognitive load. Specifically, if CL is high and learning performance decreases, ECL might likely be inhibiting successful learning. If CL is high and learning performance increases, higher mental effort might foster knowledge acquisition processing. Nevertheless, the differences between fostering and inhibiting cognitive load might show up on an individual level. That is, taking average scores in general, non-linear information planning might have a disadvantageous impact on learning performance compared to linear information presentation, but it might be compensated by a higher degree of mental effort (Hypothesis 3).

5. Method

5.1. Design and sample

In order to test these hypotheses, three factors, Text Type, Presentation Format, and Modality were combined factorially to yield eight experimental cells. The resulting design was a 2 *text type* (text with narrative structure vs. text without narrative structure) × 2 *presentation format* (linear vs. non-linear), × 2 *modality* (text vs. audio) factorial design. 125 university students sampled from a large Austrian university participated in the experiment (mean age = 24.19, SD = 5.10). The students, (87 female and 38 male), were randomly sampled from a variety of different fields of study and randomly assigned to one of the eight between-subjects groups. The students received either € 10 for their participation or a study-relevant certificate. Five experimental cells contained 15 participants and three cells contained 17.

5.2. Dependant variables

5.2.1. Learning outcomes

In order to measure the students' learning success, and to control for prior knowledge, a knowledge pre-and post-test was conducted. The test was a composite of two assessment methods—a multiple-choice test and an essay task. Both methods were designed to measure learners' knowledge and understanding of the basic concepts described in the texts. The multiple-choice test was comprised of 12 questions written at the comprehension level of Bloom's taxonomy (Bloom, 1956), each with one correct answer and three foils (see Table 1 for examples of the test questions). The MC test difficulty yielded values of 0.29 at pretest and 0.63 at posttest for the expository text, and 0.02 and 0.74 for the pre- and posttest, for the narrative, respectively. The number of multiple-choice questions answered correctly was used as one metric of learning success.

For the essay task, all participants were asked to write down anything they already knew about the topic of the learning task.

Table 1
Examples of multiple-choice test questions (correct answers in italics).

What is nucleic acid?
Carrier of protein-synthesis information
Carrier of enzyme information in the cell nucleus
Carrier of genetic information
Carrier of cell nucleus information
What is a gene?
DNA
Part of DNA
RNA
Part of RNA

There was no minimum or maximum number of words given, as well as no time constraint. The essay task was analyzed by a propositional analysis based on the suggestions of Kintsch and van Dijk (1978). Once essays were broken down into propositions, each proposition was allotted one point if it corresponded to the semantic equivalent of the propositions in the experimental text. In order to compare the two different text types, the proportional value was computed for each participant by dividing the number of the correct essay propositions participants produced, by the overall number of propositions in the corresponding learning material (for pre- and post-tests separately). With the knowledge pre-test being the same as the post-test, a testing effect might occur. But as this treatment is constant over all conditions, we assume that error is equally balanced between the two.

5.2.2. Cognitive load

Two instruments were used to measure cognitive load. First, the *Mental Effort Rating Scale (MERS;* Paas, Van Merriënboer, & Adam, 1994) was used to assess the mental effort participants actually invested in processing the learning material. Second, a slightly adapted form of the *NASA-TLX (Task Load Index)* was used. Developed by Hart and Staveland (1988), the NASA-TLX consists of five subscales each represented by a one-item self-report question (i.e. *Task requirements*, *Effort in understanding content*, *Expectation of success*, *Effort in navigation,* and *Stress*). Both measures (the MERS and NASA-TLX) consist of a 10-point rating scale, from 0 (completely disagree) to 10 (completely agree), and were used in the post-test phase only (NASA-TLX immediately followed by the MERS). An analysis of the internal consistency for the NASA-TLX yielded a *Cronbach's Alpha* of *0.46*. However, in order to increase the internal consistency we: (a) excluded one item from the analysis ("expectation of success"), (b) used the mean from the four remaining items, and (c) computed a new *Cronbach's Alpha.* The resulting Cronbach's Alpha equalled *0.71*. It is important to note that using self-reported measures of cognitive load is not unproblematic (cf. Brünken, Plass, & Leutner, 2003; de Jong, 2010); however, they are currently considered the most common, reliable, and valid approaches of cognitive load assessment (see Paas et al., 2003).

5.3. Materials

In order to analyze the influence of text type and modality, eight different learning environments were developed. The operationalization of two different content/text types that have a different process focus (acquiring single concepts and details vs. fostering holistic text processing that contributes to an understanding of the plot) was accomplished by comparing two different texts. One was a narrative text with a linear plot; the other was an expository text derived from online encyclopaedia sources on molecular biology. For the narrative text type, we chose a relatively unknown fairy tale, 3083 words in length, from Hans Christian Andersen's "The Girl Who Trod on the Loaf" (Hersholt, n.d.). The rationale for the choice of a fairy tale as a script-based text type was based on the fact that this genre typically has a chronological plot with a clear beginning and an end, in addition to several protagonists, each with clear character descriptions and interactions. Narratives are a type of text genre typically used for moral education that provides an ideal learning scenario. The expository text was a well-integrated explanatory passage, 2538 words in length, using simple accessible language to describe the composition, relationship, and function of bacteria, viruses, and genes. Both texts were divided into locally coherent nodes, with each node's content understandable in itself without the need to refer to other nodes. The narrative and expository texts were comprised of 35 and 19 nodes, respectively, and were presented in the native language of the learners.

The expository articles had a higher complexity than the narrative text. In the narrative text, there were more bridging sentences in order to guarantee local coherence. By reducing the amount of words in the expository text, we balanced task difficulty resulting from the complexity of the expository articles.

Next, each text was transformed into the computer-based learning environments with either a linear or a non-linear navigation access. In the linear condition, navigation was possible by means of a "next", "back", and "starting page" button. The sequence in the narrative text was determined by its chronological narrative structure; in the linear expository conditions, we arranged the nodes in alphabetical order depending on the topic of each node. For the non-linear hypertext conditions, we linked the nodes associatively by selecting keywords and phrases as hyperlinks that referred to other nodes explaining and elaborating these terms more deeply. In addition, we added buttons leading to the starting page and back to the last visited page. Thus, there was no way for learners, in the nonlinear condition, to move from page to page sequentially because: (a) movement was only possible by clicking on hyperlinks or by going back to the main page, and (b) the hyperlinks did not link the pages in a linear manner. In short, there was no "next" button on any page in the nonlinear condition.

The operationalization of the modality (written text vs. audio) was accomplished by using the different versions of learning material for the text condition. That is, a professional speaker recorded all written texts for auditory presentation. For presentation purposes, we simulated a cellular phone interface run as a computer program in order to integrate the recorded audio files.

Corresponding to the text-based conditions, the resulting computer programs simulated a Hyperaudio learning environment navigable by using the keys of the cellular phone. There was a basic navigation set with options "Repeat", "Back" (return to the last visited audio-node), and "Main page" (see Fig. 1). When a learner opened an audio-node, a trained professional speaker presented the information of this node as audio. Simultaneously, possible hyperlinks were presented in the display assigning key-codes to each hyperlink. By pressing the corresponding key, a new audio-node opened with the corresponding auditory information (e.g., pressing the "3"-key in Fig. 1 would have opened the audio document about viruses). In the non-linear program, audio-nodes were associatively linked by showing available hyperlinks in the display of the cellular phone; in the linear program, only a "back & forward"-navigation format was provided, realized by key commands provided by the cellular phone interface. Structure and content of the audio versions of the learning material were identical to the correspondent text-based version (see Fig. 2). Treatment time was set at 35 min and was constant over all conditions. This allowed all participants to go through all learning materials at least twice.

5.4. Procedure

The experimental procedure began with oral and written instructions about the task and the way to use the learning pro-

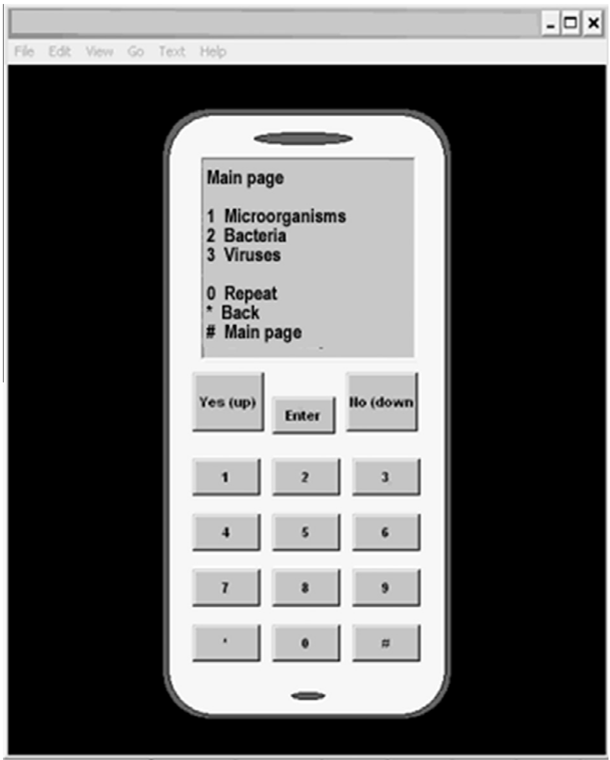

Fig. 1. Interface of the Hyperaudio learning environment application running on a cellular phone.

grams, followed by the pre-test. In the learning phase, participants had 35 min to navigate through the programs, with note-taking not allowed. The programs were delivered on single desktop PCs. Following program navigation, the participants were directed to complete the post-test with the same essay task and multiple-choice test as in the pre-test, in addition to the NASA-TLX and the MERS. The pre- and post-test lasted 30 min each. All participants fully completed participation in the experiment.

5.5. Data source

After analyzing the internal consistencies of all scales (see above), all single items were aggregated into scales using the arithmetic mean, except the scores from the multiple-choice tests where the sum values were computed. All inferential statistical testing were accomplished with a Multiple Analysis of Covariance (MANCOVA) using the statistical software package PASW 18. Since the pre-test values in mental effort and the measures from both knowledge tests were positively correlated with the corresponding post-test values, all the pre-test variables were entered into the analysis as covariates. In short, "Presentation Format", "Modality", and "Text Type" were set as fixed factors, and all other variables were entered into the analysis as dependent variables.

6. Results

6.1. Influence of modality

Our first question was aimed at whether learners in the auditory condition would show better learning performance in measures of overall text comprehension. The results revealed that the text condition led to better performance in both knowledge assessments, in addition to higher values in self-reported cognitive load, relative to the audio condition (see Table 2). The analysis yielded an overall effect on all dependant variables ($F(3,113) = 24.74$, $p < 0.001$, $\eta^2 = 0.39$). While modality had a marginally significant impact on the MCT-post-test ($F(1,115) = 3.12$, $p = 0.08$, $\eta^2 = 0.03$), participants in the audio condition scored significantly lower in the Essay-Task-post-test ($F(1,115) = 63.53$, $p < 0.001$, $\eta^2 = 0.36$); participants also reported a higher Cognitive Load as assessed with the NASA-TLX ($F(1,115) = 7.57$, $p = 0.007$, $\eta^2 = 0.06$).

6.2. Linear vs. non-Linear information retrieval

Our second question addressed the impact of linear vs. non-linear information access. Using the same design as above, the analysis yielded an overall effect of this factor on all dependant variables

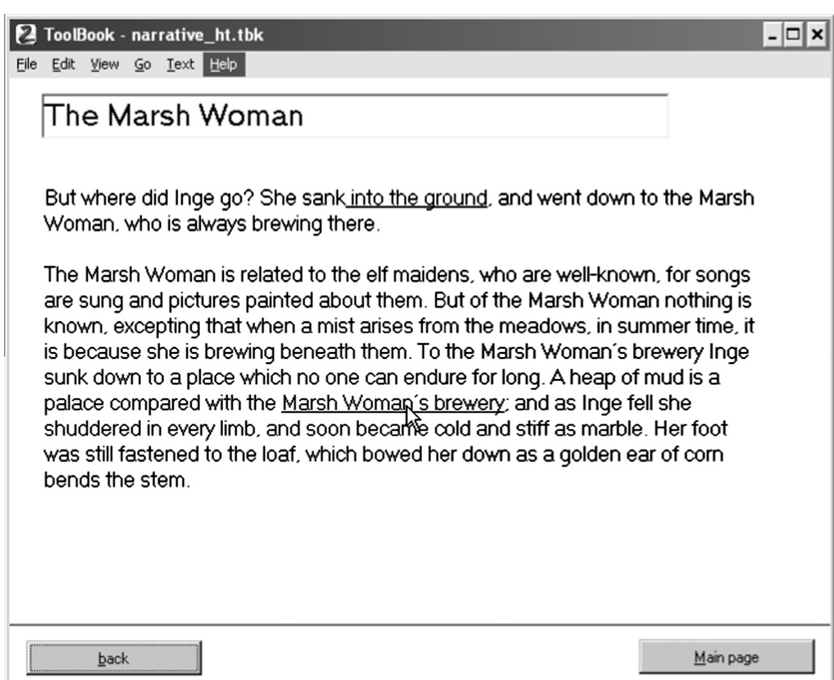

Fig. 2. Sample screen from the text-based non-linear condition.

($F(3,113) = 3.92$, $p < 0.01$, $\eta^2 = 0.09$). Participants using the site by linear navigation performed significantly better in the MCT-post-test than participants accessing information in a non-linear fashion ($F(1,115) = 4.59$, $p = 0.03$, $\eta^2 = 0.04$); and, learners in the non-linear condition reported a significantly higher level of cognitive load ($F(1,115) = 6.14$, $p = 0.02$, $\eta^2 = 0.05$). There was no significant difference in performance on the Essay Task post-test ($F(1,115) = 2.80$, $p = 0.10$, $\eta^2 = 0.02$).

6.3. Text type

Our third questions were concerned with the impact of Text Type. We hypothesized that auditory information access would lead to decreased learning performance compared to text-based non-linear learning environments when learning in a domain with a focus on single and unrelated concepts. This effect was expected to vanish with a text that followed a strong plot. Outcomes of the MANCOVA revealed a significant main effect for text type ($F(3,113) = 10.77$, $p < 0.001$, $\eta^2 = 0.22$). Specifically, participants in the expository text type condition showed scores significantly lower in both knowledge post-tests (MCT-post-test: $F(1,115) = 22.90$, $p < 0.001$, $\eta^2 = 0.17$, and Essay Task post-tests: $F(1,115) = 19.27$, $p < 0.001$, $\eta^2 = 0.14$). There was no significant effect for Cognitive Load ($F(1,115) = 0.14$, $p = 0.71$, $\eta^2 = 0.001$).

6.4. Interactions of factors

The two-way interaction of Presentation Format × Text Type was statistically significant ($F(3,113) = 4.13$, $p < 0.008$, $\eta^2 = 0.10$). Outcomes indicated that there were small differences between the linear and non-linear presentations of the expository text type on both measures of knowledge acquisition (MCT-post-test: $F(1, 15) = 5.48$, $p = 0.02$, $\eta^2 = 0.05$; Essay Task-post-test: $F(1,115) = 6.46$, $p = 0.01$, $\eta^2 = 0.05$), as well as Cognitive Load, ($F(1,115) = 3.59$, $p = 0.06$, $\eta^2 = 0.03$. Presenting the linear text type in the non-linear manner led to decreased performance in the knowledge post-tests (see Fig. 3), and increased Cognitive Load (see Fig. 4).

There was also a significant interaction of Modality × Text Type ($F(3,113) = 4.18$, $p < 0.008$, $\eta^2 = 0.10$). The linear text type led to better learning outcomes when presented as written text than as expository text. These differences were reduced when presented as audio, although performance in the audio condition was still worse than the visual presentations (as assessed with the Essay-Task-post-test: $F(1,115) = 8.80$, $p = 0.004$, $\eta^2 = 0.07$ (see Table 2). Cognitive load failed to yield a reliable influence on these measures ($F(1,115) = 2.18$, $p = 0.14$, $\eta^2 = 0.02$) or MCT-post-test scores ($F(1,115) = 0.48$, $p = 0.49$, $\eta^2 = 0.004$).

Another interaction effect was found for Presentation Format × Modality ($F(3,113) = 2.87$, $p < 0.04$, $\eta^2 = 0.07$). There were small differences between audio and text in the non-linear conditions; in the linear condition, however, the difference in Cognitive Load increased between audio and text, with text having far higher cognitive load than audio ($F(1,115) = 5.37$, $p = 0.02$, $\eta^2 = 0.05$. There was no effect on both measures of knowledge acquisition (Essay Task-post-test: $F(1,115) = 1.52$, $p = 0.22$, $\eta^2 = 0.001$; MCT-post-test: $F(1,115) = 0.36$, $p = 0.55$, $\eta^2 = 0.003$; see Fig. 5).

The three-way interaction was not significant ($F(3,113) = 1.78$, $p = 0.15$, $\eta^2 = 0.05$).

Our final research question addressed the influence of prior knowledge and mental effort within the treatment conditions. We assumed that non-linearly presented information might have a disadvantageous impact on learning performance compared to linearly presented information, although the former might be compensated by a higher degree of mental effort. The results revealed overall effects for all three covariates (MCT-pre-test: $F(3,113) = 3.91$, $p = 0.01$, $\eta^2 = 0.09$; Essay Task-pre-test: $F(3,113) = 2.68$, $p = 0.05$; $\eta^2 = 0.07$; and Mental Effort: $F(3,113) = 18.24$, $p < 0.001$, $\eta^2 = 0.33$). Tests for the specific effects revealed an influence of the covariate MCT pre-test on the performance in the MCT post-test ($F(1,115) = 7.19$, $p = 0.008$, $\eta^2 = 0.06$), but not on the other dependent variables (Essay Task-post-test: $F(1,115) = 0.81$, $p = 0.37$, $\eta^2 = 0.007$, or cognitive load: ($F(1,115) = 2.45$, $p = 0.12$, $\eta^2 = 0.02$). The covariate Essay task-pre-test had a significant influence on the performance in the MCT post-test ($F(1,115) = 4.99$, $p = 0.03$, $\eta^2 = 0.04$) and in the Essay Task post test ($F(1, 115)$=5.43, $p = 0.02$, $\eta^2 = 0.05$), but not on Cognitive Load ($F(1,115) = 0.16$, $p = 0.69$, $\eta^2 = 0.001$). Mental effort had its only impact on Cognitive Load ($F(1,115) = 50.86$, $p < 0.001$, $\eta^2 = 0.31$), while both other dependent variables missed significant F-values (MCT-post-test: $F(1,115) = 3.45$, $p = 0.07$, $\eta^2 = 0.03$ and Essay Task-post-test: $F(1,115) = 1.05$, $p = 0.31$, $\eta^2 = 0.009$).

6.5. Analyzes of correlations

Pearson correlations indicated moderate relationships between MCT-pre-test results and Cognitive Load ($r = 0.22$, $p < 0.05$), Mental Effort ($r = 0.24$, $p < 0.01$), the Essay Task post-test ($r = -0.38$, $p < 0.01$), and the MCT post-test performance ($r = 0.64$, $p < 0.01$). Similar correlations were observed for Essay Task pre-test with Cognitive Load ($r = 0.24$, $p < 0.01$) and Mental Effort ($r = 0.27$, $p < 0.01$). Mental Effort correlated significantly with the MCT post-test ($r = -0.34$, $p < 0.01$), cognitive load ($r = 0.60$, $p < 0.01$) and the Essay Task post-test ($r = -0.27$, $p < 0.01$).

6.6. Regression analysis

A regression analysis using all covariates as independent variables revealed only Mental Effort as a significant predictor for performance on the MCT post-test ($F(1,125) = 16.62$, $p < 0.001$) with the regression equation yielding MCT post-test = 11.47 + (−0.71) × Mental Effort ($R^2 = 0.12$). A second regression analysis using all covariates as independent variables revealed Mental Effort, MCT-pre-test, and Essay Task-pre-test performance as significant predictors for performance in the Essay Task post-test score ($F(1,125) = 9.92$, $p < 0.001$), with the regression equation yielding Essay Task post-test = 0.81 + (−0.13) × Mental Effort + (−0.13) × MCT-pre-test + 1.84 × Essay Task-pre-test ($R^2 = 0.20$).

7. Summary and discussion

The results revealed that presenting information in a linear manner leads to better learning outcomes overall than information presented non-linearly. The main effect of Presentation Format showed that, in general, the non-linear information presentation led to an increase in cognitive load. This increase is in line with the assumption that navigation planning in non-linear learning environments might be a secondary task that increases extraneous cognitive load (like in text-hypertext comparisons; cf. Wells & McCrory, 2011). Nevertheless, linearity apparently has only a small impact on the performance of a fact-based multiple-choice test and no impact on the mental representation of the text structure and its plot as assessed by the essay task. This finding is in line with prior research findings showing comparable outcomes between learning with text and Hypertext (Chen & Rada, 1996; Shapiro & Niederhauser, 2004).

The findings also reveal that presenting learning material as written text is more effective than presenting the learning material in an auditory format. Presenting information as written text leads to increased performance in knowledge acquisition as measured in

Table 2
Mean values and standard deviations.

Measure	Presentation format	Text type	Modality	Mean	SD
MC-pre-test	Linear	Expository	Text	3.47	1.46
			Audio	3.24	2.40
		Narrative	Text	0.00	0.00
			Audio	0.76	1.03
	Non-linear	Expository	Text	3.3	1.40
			Audio	4.07	2.28
		Narrative	Text	0.00	0.00
			Audio	0.29	0.77
MC-post-test	Linear	Expository	Text	8.07	1.98
			Audio	6.59	2.90
		Narrative	Text	9.53	1.51
			Audio	9.94	1.71
	Non-linear	Expository	Text	8.00	1.65
			Audio	7.47	1.85
		Narrative	Text	8.33	1.63
			Audio	7.59	2.00
Essay task-pre-test	Linear	Expository	Text	0.012	0.009
			Audio	0.006	0.006
		Narrative	Text	0.000	0.000
			Audio	0.000	0.000
	Non-linear	Expository	Text	0.013	0.006
			Audio	0.011	0.009
		Narrative	Text	0.000	0.000
			Audio	0.000	0.000
Essay task-post-test	Linear	Expository	Text	0.10	0.03
			Audio	0.05	0.03
		Narrative	Text	0.21	0.07
			Audio	0.10	0.06
	Non-Linear	Expository	Text	0.11	0.03
			Audio	0.06	0.03
		Narrative	Text	0.15	0.04
			Audio	0.09	0.05
Cognitive load (NASA-TLX)	Linear	Expository	Text	6.94	1.15
			Audio	5.13	1.43
		Narrative	Text	4.51	2.38
			Audio	3.76	2.00
	Non-linear	Expository	Text	6.51	1.47
			Audio	6.07	1.38
		Narrative	Text	5.84	1.87
			Audio	5.93	1.83

both learning measures. This is, at a first glance, a contradictory finding relative to prior research in reading vs. listening comprehension (e.g., Golas, Orr, & Yao, 1994). But, in the present investigation, all text conditions yielded cognitive load significantly higher than in the audio conditions of the learning material. Certainly, this finding seems to be contradictory because higher extraneous cognitive load might lead to inhibited knowledge acquisition. However, we contend that the NASA-TLX does not distinguish between an inhibition of extraneous cognitive load and mental effort that might contribute to elaboration processes. If this is true, it is reasonable to expect a rather high correlation between the NASA-TLX and the Mental Effort Rating Scale as well as a negative correlation with Mental Effort on post-test-performance in knowledge assessment measures. Indeed, the pattern of these correlations was observed in the regression analyzes. It is also particularly noteworthy that, in the essay task, regression analysis and the influence of the covariates indicated that only prior knowledge was a strong predictor for performance in the knowledge post-test. Thus, the lack of differentiation between a positive and negative influence of cognitive load on the instruments used here leads to an interpretation of a modality effect that can be attributed to a modality- or media-specific appraisal effect resulting in different mental effort. Comparable to the study conducted by Salomon (1984), Salomon and Leigh (1984), participants might have judged the auditory learning environments as rather easy and, thus, might have invested less mental effort while participants in the text-based conditions might have encountered their learning environments as rather difficult and, thus invested more mental effort to process the information to them. Additional theoretical support for this interpretation is provided by Geary (2004). He differentiates between biologically primary and secondary knowledge. Processing auditory information is a primary skill where little working memory is used to process this kind information; therefore, learners experience the learning process as easy. Reading, by comparison, is a biologically secondary skill, likely to be experienced as more difficult. Thus, the more learners experience the task as difficult, the more likely they are to expend more mental effort and the more working memory resources they must activate to accomplish the task.

On the other hand, we found no evidence that differences between auditory and text-based instruction supports either concept-related knowledge or construction of a global coherent structure of the text. The low correlation among the measures used here (Multiple-Choice Test and Essay Task) indicates that the measures present two different and valid approaches to assess the influence of the experimental variables. Nevertheless, a possible explanation for the superiority of text-based instruction might be a lack of invested mental effort in processing the auditory information, as discussed above.

Finally, the poorer performance of the auditory conditions compared to the written text might also be explained by the transient information effect (cf. Leahy & Sweller, 2011). This effect refers to

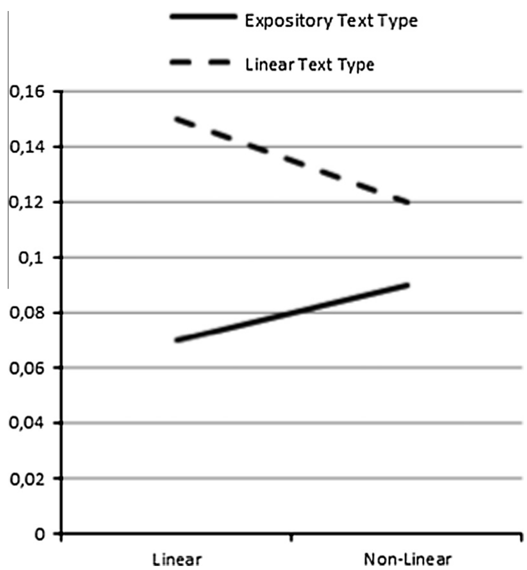

Fig. 3. Interaction between presentation format and text type on essay-task-post-test-performance.

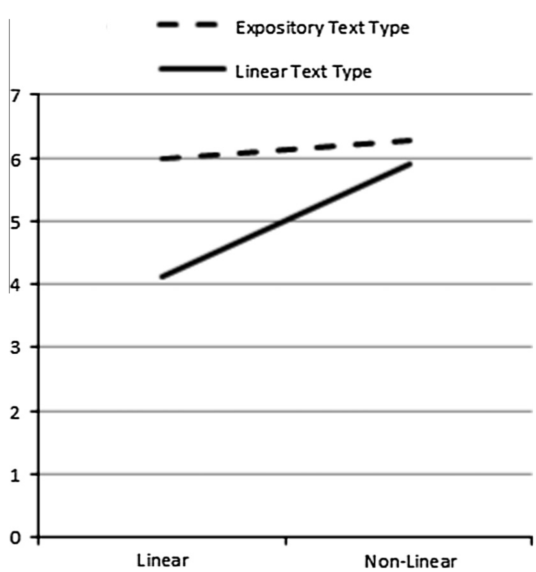

Fig. 4. Interaction between presentation format and text type on cognitive load.

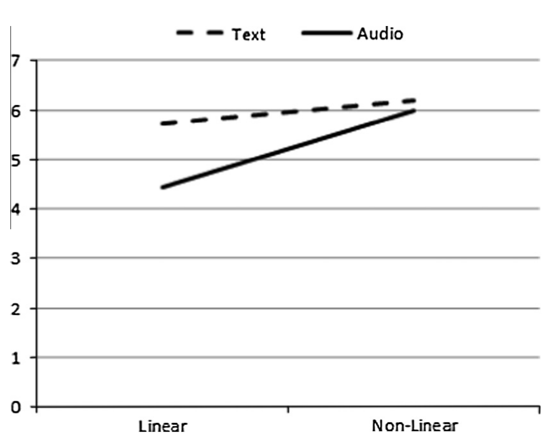

Fig. 5. Interaction between presentation format and modality on cognitive load.

the observation that longer and complex sentences presented verbally might not be processed because they exceed working memory capacity. In contrast, visually presented text with the same length and complexity can be steadily rescanned and, thus, processed without information loss. This would also support the assumption that the differences between reading and listening in this study do not originate from modality specific information processing (cf. Rummer et al., 2008), but rather from different rehearsal mechanisms (see also Kürschner et al., 2006).

Another major aim of this research was to analyze specific interaction effects that might influence learning outcomes. The interaction between Modality and Presentation Format indicates that additional mental effort necessary for deeper processing is applied in linear conditions; by comparison, in the non-linear conditions, a possible ceiling effect levels out the difference because extraneous cognitive load is caused by additional navigation planning.

Results of this study unambiguously suggest that different representations (text vs. audio and linear vs. non-linear) are affected by the nature of the learning material. The interaction effect between Text Type and Presentation Format reveals that the structure of the text and, thus, the structure of the content within the two presentation contexts have a direct impact on cognitive load and learning outcomes. This is consistent with research presented by Zumbach and Mohraz (2008): a primarily linear structured text with an underlying linear plot presented in a non-linear manner leads to increased cognitive load. We interpret load as an increase in *extraneous* cognitive load because reconstructing the linear plot is an additional secondary task that is not supported by the non-linear structure of the learning environment. This assumption is supported by performance measures in the knowledge assessments: the linear text type presented either as Hypertext or Hyperaudio led to poorer performance in the multiple-choice test as well as the essay task. Likewise, the format of presentation (linear vs. non-linear) also affected the rather loosely structured expository text type. The two-way-interactions support the assumption that there are slight advantages of presenting this kind of content in a non-linear learning environment although the effect is smaller compared to the effect of presenting linear text in a non-linear manner.

It is important to point out that the interaction effect discussed above could also have been fostered by the difficulty of the learning material. The main effect shows that, in both knowledge assessments, the expository text type led to poorer performance than the linear text type. This suggests that the content was much more difficult to understand than the linear text and, thus, might have increased intrinsic cognitive load. However, it did not affect the self-reported cognitive load. Thus, the origin of the interaction effects can be best attributed to the structure of the content rather than to the difficulty. In addition, comparability of both text types was observed not only by comparing fact-related knowledge, but also by using the standardized amount of newly acquired propositions by means of the essay task.

Taken together, the results show that when presenting information in a linear or non-linear manner, the nature of the content and, thus, the structure of the text have to be regarded. This is a factor commonly neglected in most studies comparing linear with non-linear media. Going back to the first instructional approaches on hypermedia learning and to cognitive flexibility theory (e.g., Jacobson & Spiro, 1995; Spiro & Jehng, 1990), we have to re-interpret findings from early evidence that non-linear information retrieval contributes directly to cognitive flexibility: When Spiro and Jehng (1990) used hypertext in their studies as a learning environment for movie interpretation, we have to assume that participants already had a mental representation about the movie and its plot. Thus, the participants in Spiro and Jehng's study did not have to reconstruct the linear sequence from a non-linear learning environment; rather, they were able to re-use their existing knowledge for constructing the hypertext information into a mental representation of the movie plot AND its interpretation. This stands in line with most hypermedia research findings: that prior knowledge is a main predictor for successful learning with hypermedia. Nevertheless, if this prior knowledge is not available, the assembly of content structure and representational structure (likewise linear vs. non-linear) has to be a crucial instructional design task.

Finally, findings from the present experiment regarding auditory instruction suggest that additional support might be necessary for enhancing learning from linear audio as well as Hyperaudio, e.g., for learning with mobile or handheld computing devices.

The implications from this experiment reveal that using Hyperaudio, as a stand-alone instructional device, is disadvantageous compared to the other representation formats used in this study. Nevertheless, the outcomes of this research also suggest that the dominating research practice in media comparison might be problematic: that is, it does not necessarily matter in which modalities, or more precisely codalities, information is presented. Rather, it is more important to take the structure and the content of the underlying information into account. Here the interaction effects between text type and the presentation format were not only possible, but rather likely.

With the introduced concept of "Hyperaudio", we have to conclude that this instructional design either needs support by scaffolds designed to enhance learning or has to be carefully integrated into broader learning environments. Also additional instructions might be helpful and necessary to increase mental effort and, thus, elaborate information. As this re-search on non-linear auditory instruction and its support is still emerging, further research from a technical, content-specific didactical, and a psychological perspective, is needed.

References

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.

Baddeley, A. D. (1998). *Human memory; theory and practice*. Boston: Allyn and Bacon.

Barron, A. E. (2004). Auditory instruction. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 949–978). Mahwah, NJ: Lawrence Erlbaum.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives handbook 1: Cognitive domain*. New York, NY: Longman, Green & Co.

Brünken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53–61.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human–Computer Interaction, 11*, 125–156.

Conklin, J. (1997). Hypertext: An introduction and survey. *IEEE Computer, 20*(9), 17–41.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105–134.

Folz, P. W. (1996). Comprehension, coherence, and strategies in hypertext. In J. F. Rouet (Ed.), *Hypertext and cognition* (pp. 109–136). Mahwah, NJ: Erlbaum.

Geary, D. C. (2004). Evolution and cognitive development. In R. Burgess & K. MacDonald (Eds.), *Evolutionary perspectives on human development* (pp. 99–133). Thousand Oaks, CA: Sage Publications.

Golas, K. C., Orr, K. L., & Yao, K. (1994). Storyboard development for interactive multimedia learning. *Journal of Interactive Instruction Development, 6*, 18–31.

Grabinger, S., & Dunlap, J. C. (1996). Nodes and organization. In P. A. M. Kommers, S. Grabinger, & J. C. Dunlap (Eds.), *Hypermedia learning environments* (pp. 79–114). Mahwah, NJ: Lawrence Erlbaum.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North Holland.

Hartley, J. (2004). Designing instructional and informational text. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 917–947). Mahwah, NJ: Lawrence Erlbaum.

Hersholt, J. (n.d.). *The Girl Who Trod on the Loaf*. Online Document retrieved from <http://www.andersen.sdu.dk/vaerk/hersholt/TheGirlWhoTrodOnTheLoaf_e.html>. Date of access: 05.11.12.

Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes, 1*, 91–117.

Hsi, S. (2003). A study of user experiences mediated by nomadic web content in a museum. *Journal of Computer Assisted Learning, 19*(3), 308–319.

Jacobson, M. J., & Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research, 12*(5), 301–333.

Johnson-Laird, P. N. (1983). *Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363–394.

Kürschner, C., & Schnotz, W. (2008). Das Verhältnis gesprochner und geschirebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau, 59*(3), 139–149.

Kürschner, C., Schnotz, W., & Eid, M. (2006). Konstruktion mentaler Repräsentationen beim Hör- und Leseverstehen. *Zeitschrift für Medienpsychologie, 18*(2), 48–59.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943–951.

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human–computer relationship*. Cambridge, MA: The MIT Press.

Nass, C., & Lee, K. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity–attraction, and consistency–attraction. *Journal of Experimental Psychology: Applied, 7*, 171–181.

Nugent, G. C. (1982). Picture, audio, and print: Symbolic representations and effect on learning. *Educational Communications and Technology, 30*, 163–174.

O'Bannon, W., Lubke, J., Beard, J., & Britt, V. (2011). Using podcasts to replace lecture: Effects on student achievement. *Computers & Education, 57*, 1885–1892.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.

Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual Motor and Skills, 79*, 419–430.

Paechter, M. (1996). *Auditive und visuelle Texte in Lernsoftware*. Münster: Waxmann.

Roschelle, J., & Pea, R. (2002). A walk on the WILD side: How wireless handhelds may change CSCL. *International Journal of Cognition and Technology, 1*(1), 145–168.

Rummer, R., Schweppe, J., Scheiter, K., & Gerjets, P. (2008). Lernen mit Multimedia: Die kognitiven Grundlagen des Modalitätseffekts. *Psychologische Rundschau, 59*, 98–107.

Salmerón, L., Baccino, T., Cañas, J. J., Madrid, R. I., & Fajardo, I. (2009). Do graphical overviews facilitate or hinder comprehension in hypertext? *Computers & Education, 53*, 1308–1319.

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attribution. *Journal of Educational Psychology, 76*(4), 647–658.

Salomon, G., & Leigh, T. (1984). Predispositions about learning from print and television. *Journal of Communication, 32*, 119–135.

Schnotz, W., & Zink, T. (1997). Informationssuche und Kohärenzbildung beim Wissenserwerb mit Hypertext. *Zeitschrift für Pädagogische Psychologie, 11*(2), 95–108.

Shapiro, A., & Niederhauser, D. (2004). Learning from hypertext: Research issues and findings. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 605–620). Mahwah, NJ: Lawrence Erlbaum Associates.

Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163–205). NJ: Erlbaum; Hillsdale.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257–287.

Vajoczki, S., Watt, S., Marquis, N., & Holshausen, K. (2010). Podcasts: Are they an effective tool to enhance student learning? A case study. *Journal of Educational Multimedia and Hypermedia, 19*, 349–362.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Orlando: Academic Press Inc.

Wells, A. T., & McCrory, R. (2011). Hypermedia and learning: Contrasting interfaces to hypermedia systems. *Computers in Human Behavior, 27*, 195–202.

Zahn, C., Schwan, S., & Barquero, B. (2002). Authoring hypervideos: Design for learning and learning by design. In R. Bromme & E. Stahl (Eds.), *Writing hypertext and learning* (pp. 153–176). Amsterdam: Pergamon.

Zumbach, J. (2006). Cognitive overhead in hypertext learning re-examined: Overcoming the Myths. *Journal of Educational Multimedia and Hypermedia, 15*(4), 411–432.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior, 24*(3), 875–887.