

## COMPREHENDING NEWS VIDEOTEXTS: THE INFLUENCE OF THE VISUAL CONTENT

**Jeremy Cross**

**Nanyang Technological University**

Informed by dual coding theory, this study explores the role of the visual content in L2 listeners' comprehension of news videotexts. L1 research into the visual characteristics and comprehension of news videotexts is outlined, subsequently informing the quantitative analysis of audiovisual correspondence in the news videotexts used. In each of five lessons, ten pairs of Japanese EFL learners participated in a sequence of tasks in which they listened to, and discussed various facets of their comprehension of news videotexts. The pairs' dialogue acted as the unit of analysis for exploring the effect of visual information on their comprehension. The qualitative analysis illustrated that various attributes of the visual content, such as audiovisual correspondence, impacted on comprehension. Moreover, other influences of the visual content found were its general utility in facilitating comprehension, inhibiting of attention to, and processing of audio information, and stimulation of learners' expectations and inferencing of content. Based on these findings, learner variability aspects and several implications for related L2 listening pedagogy are discussed.

### INTRODUCTION

Advances in satellite, digital video and broadband technology mean that news videotext services are readily available to viewers across the globe. L2 users form a large part of the world-wide audience, with news videotexts providing them with an authentic sociocultural, linguistic and educational resource which can be exploited for language learning inside and outside the classroom. However, the intrinsic audiovisual nature of news videotexts means that L2 users not only have to deal with the challenges to listening comprehension<sup>1</sup> that they typically encounter which are associated with the audio channel (e.g., unfamiliar vocabulary, speech rates, prosody and syntactic structures), but also need to cope with the vagaries of content presented in the accompanying visual channel if they are to process, understand, and respond to the message news videotexts are crafted to convey. A number of publications point to the correspondence between audio and visual information as one potentially important factor affecting L2 learners' comprehension (Meinhof, 1994, 1998). However, while such intuitions regarding the influence of visual elements seem valid, there is very little empirical research which is informative in such respects. Moreover, apart from Gruba's (2004, 2006) studies, little is known about how L2 listeners strategically exploit visual content in news videotexts to facilitate comprehension. Given that the use of news videotexts in second and foreign language classrooms and self-access centres is increasingly common practice (particularly with more advanced listeners), there is a need for related research which promotes understanding of the influence of the visual content in L2 listeners' comprehension of this videotext genre. This paper reports on a study which draws on relevant L1 and L2 theory and empirical research to address this issue.

### BACKGROUND

#### Audiovisual Processing

In line with L1 research into the processing of audiovisual information in multimedia (Mayer & Anderson, 1991; Mayer & Sims, 1994) and in news videotexts in particular (Walma van der Molen & Van der Voort, 2000; Walma van der Molen, 2001), the theoretical perspective underpinning this study is dual coding theory. In his seminal work, Paivio (1971, 1990, 2007) proposes dual coding theory as a

theory of cognition, which is distinguished from other common-coding theories of cognition (e.g., propositional representation) by its modality-specific nature. That is, it provides a coherent account of how *separate* verbal and nonverbal mental representations are collectively processed. The basic premises of dual coding theory most recently presented by Paivio (2007), which builds on his own early work and also research in association with several colleagues (Clark & Paivio, 1991; Sadoski & Paivio, 2001), are:

- Both verbal and nonverbal systems are specialized and distinct, and mental representations associated with each system preserve the properties of the sensorimotor events which trigger them;
- The verbal system encompasses written, auditory, and articulatory verbal codes;
- The nonverbal system includes images for environmental sounds, activities, and events;
- While written, aural and articulatory input is each typically processed sequentially by the verbal system, the nonverbal system processes information simultaneously as a whole, for a single mental image comprises a multitude of details;
- The verbal and nonverbal systems are joined by referential connections as part of a complex associative network (e.g., imagery may evoke word representations and vice versa);
- Associative connections are another type of link within each of the verbal and non verbal systems (e.g., a word or an image may activate associated words or images, to create complex configurations of mental representations);
- The activation of mental representations in either system may or may not be a conscious experience;
- Patterns of connection activation are influenced by contextual factors (e.g., a particular task such as showing pictures may prime the nonverbal system and promote the production of mental images);
- Verbal and nonverbal mental representations and their interconnections differ for each individual due to their diverse past experiences;
- And, nonverbal processing is affected by an individual's propensity and capability to use imagery.

As a hypothetical example of dual coding theory in the context of this study, a visual scene/shot and its accompanying audio content in a news videotext would activate, depending on the individual's capacity, corresponding mental representations in both verbal and nonverbal systems, some of which are conscious. Spreading activation through associative and referential connections would occur within and between the two systems generating an intricate and idiosyncratic pattern of mental representations which need to be filtered to formulate a correct interpretation. Extending this hypothesis further, verbal and imagery representations activated by complementary stimuli would potentially generate relatively less complex mental patterns than when incongruence is evident, with associated positive and negative consequences for cognitive loading, respectively.

### **Videotexts**

A videotext is broadly defined here as a multimodal text consisting of contiguous, dynamic, and interwoven sounds (verbal, musical and/or background) and visual images (still, moving, text and/or graphic) which can be presented using a range of media. Movies, game and talk shows, dramas, music videos, documentaries, and news are all prevalent genres of videotexts, and are representative of the multitude of such material which is accessible around the world in many languages, through both satellite and terrestrial television and the Internet, to an increasingly visually-oriented populace (Meinhof, 1998).

Videotext genres differ in the extent to which they aim to entertain and/or inform an audience. Broadly speaking, for example, movies and music videos are primarily entertainment focused, whereas documentaries and news are essentially purveyors of factual information. Videotexts also vary in their degree of structure. For instance, movies and music videos are at the less-structured end of the continuum, while news, talk shows, and soap operas are notably tightly structured in contrast (Meinhof, 1998). Also, while there are format similarities in the more-structured videotexts mentioned, their production and construction reflects the sociocultural values and norms of the country or region from which they emanate (Meinhof, 1998).

In terms of language teaching and learning, the exploitation of videotexts is commonplace. Reasons for using videotexts are that the visual channel provides learners with opportunities to see and hear the target language in use and shows many aspects (e.g., landscapes, locations, fashion, food, gestures, way of life) of the target culture and society, both of which can raise learners' interest levels (Harmer, 2001; Sherman, 2003). In addition, videotexts have ecological validity, as learners are highly likely to listen to another language through this multimodal medium (Guichon & McLornan, 2008).

As with other major genres of videotexts, news videotexts (both authentic and non-authentic) are a valuable and widely used resource for advancing language learners' listening abilities, and a growing number of publications continue to offer suggestions for exploiting this material in the classroom (Gruba, 2005; Harmer, 2001; Lynch, 2009; Meinhof, 1998; Sherman, 2003). Nonetheless, despite the utility and prevalence of news videotexts in foreign and second language learning contexts, it is only recently that L2 researchers have again, following an early study by Brinton and Gaskill (1978), begun to empirically investigate ways to facilitate learners' comprehension of this genre (Cross, 2009; Rovens Mompean & Guichon, 2009). As yet, however, little research has concentrated on understanding the influence on L2 learners' news videotext comprehension of the associated visual content, which is a central element of the message this genre is fashioned to communicate, but one which is often dismissed as less important than the aural content (Graddol, 1994). Prior to exploring related L2 research, L1 research informative to this study is presented.

## **The Visual Content of News Videotexts**

### ***L1 Research***

In one of the first publications to cover news videotext comprehension, Gunter (1987) states that the reasons behind inserting visual content (rather than just including a newscaster) in news production are that it increases the overall impact of the news broadcast, serves to emphasize specific aspects of the narrative (e.g., who was involved and where the story occurred), gives the audience the impression they are being allowed to witness the reported events first-hand as they unfold, and triggers an emotional reaction. Gunter (1987) reviewed early studies from the 1960s to 1980s on the influence of the visual channel. Generally, findings from the studies presented did not offer conclusive support for visual content in enhancing information assimilation and retention in news videotexts, but the degree of redundancy between the content of the two channels was identified as an important variable in information processing.

In subsequent research, Brosius, Donsbach, and Birk (1996) suggest that the visual content of news videotexts largely consists of 'standard scenes', that is, shots of buildings, shoppers strolling in the street, or employees at work, which typically carry little information, and merely have a thematic correspondence with the audio content. Brosius, et al. (1996) investigated the effect of such standard scenes on the quality of information recalled by L1 users, representing how well the content had been conveyed, compared to three other conditions: (a) audiovisual correspondence, (b) audio content only, and (c) audiovisual divergence. The researchers found that the uptake of information was highest for audiovisual correspondence, followed by standard scenes and audio only, which both had similar recall quality. Audiovisual divergence hindered uptake the most. This outcome highlights that visual content has

the potential to facilitate news videotext comprehension when it is convergent with the audio information (see also Reese, 1984; Walma van der Molen & Van der Voort, 2000), but seems to be detrimental when there is some degree of divergence.




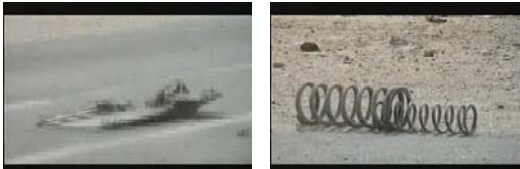

A more recent L1 study by Walma van der Molen (2001) also considered audiovisual correspondence in terms of introducing and applying a coding scheme to enable a more systematic analysis of this attribute than provided by general judgments of correspondence across whole news videotexts in previous studies. Walma van der Molen evaluated audiovisual correspondence of information presented in 'shots', that is, the visual content between 'edits' (a change to a similar scene) and 'cuts' (a change to a different visual scene), within news videotexts. Informed by earlier related research (Brosius et al., 1996; Lang, 1995), she developed four categories to establish and code the degree of semantic overlap between audio and visual channels in shots. Three of these categories are classified on a correspondence continuum ranging from Direct, through Indirect, to Divergent. In accordance with Walma van der Molen (2001), the Direct category is used to classify audio and visual content which both express the same propositional meaning (i.e., information in the two modalities is essentially semantically redundant); the Indirect category is used to classify audio and visual content which is only partly related (as in standard scenes); and the Divergent category is used to classify audio and visual content which is not related or even contradictory. The fourth category, Talking head, refers to a scene in which typically only the top half of a newsreader, reporter, or interviewee is shown as they speak, and is considered a separate category as it neither reflects conflicting audio and visual content, nor transparent semantic relatedness between the two. However, a Talking head is categorized as one of the other three categories when additional visual information is available in the background, for example, behind an interviewee. Examples of each category from a BBC news videotext about UK forces in Iraq entitled Basra Deaths are presented in [Table 1](#) for clarification.

Walma van der Molen (2001) utilized her coding system to good effect in examining a sample of Dutch news videotexts. She summarized in writing the visual content of shots and the concurrent verbal content, and noted the duration in seconds for each shot. Two coders then used the four-category scheme to code audiovisual correspondence using the written information as well as the actual news videotexts. Walma van der Molen reported a very strong inter-rater reliability (Cohen's Kappa was .81) for the two coders, which suggests her coding system is a valid and reliable method for determining the level of audiovisual correspondence in news videotexts. Her taxonomy was used in this study for the analysis of the news videotexts.

Other related L1 news videotext research has focused on graphics, such as computer-generated texts (CGTs) and computer-generated animations (CGAs). These are features which are typically used to present numerical details, and are utilized to facilitate understanding of complicated events or processes. Fox, Lang, Chung, Lee, Schwartz, and Potter (2003) investigated the comparative amount of comprehension for seven science-related news videotexts in three modified versions which contained a CGA, a CGT, or no graphics. The researchers reported that comprehension was worst for the no graphics version, though there was little difference between the CGA and CGT versions. Moreover, when the perceived complexity of the news videotext was included as a factor, comprehension was not affected for easier or harder content by the presence of a CGA or CGT, but more difficult content resulted in significantly less comprehension when no graphics were included.

In summary, this discussion illustrates that variations in audiovisual correspondence can impact on the processing of news videotexts by skilled L1 users. For L2 listeners, it seems safe to hypothesize that such factors will also be influential, as well as be potentially compounded due to linguistic deficiencies, working memory constraints, and a lack of familiarity with the culture-bound visual content, style, and conventions of news videotexts geared to the L1 audience (Meinhof, 1994).

Table 1. *Examples of the Four Coding Categories in Basra Deaths*

Verbal content	Visual content	
<p style="text-align: center;">Talking head</p> <p>The death of two British soldiers in a roadside bombing in Iraq has raised further questions about the level of equipment used by British troops and whether it's enough to fight insurgents with their increasingly sophisticated weaponry. From Iraq, David Lauren reports now.</p>		
<p>This Land Rover was hit...</p>	<p style="text-align: center;">Direct</p> 	
<p>...as British troops were escorting construction workers north of Basra this morning.</p>	<p style="text-align: center;">Divergent</p> 	
<p>Wreckage was strewn across the road...</p>	<p style="text-align: center;">Direct</p> 	
<p>...in an area where there have been similar attacks before.</p>	<p style="text-align: center;">Indirect</p> 	

**L2 Research**

Regarding L2 learning, there is a growing body of research which has investigated the influence of the visual content in videotexts primarily in terms of the role of kinesic cues (e.g., hand gestures and lip movements), and still images in lectures, and to a lesser degree dialogues, in an academic listening context (see Ginther, 2002; Ockey, 2007; Sueyoshi & Hardison, 2005; Wagner, 2007, 2008, 2010a, 2010b). The general findings of these studies were that kinesic and contextual visual cues appeared to either facilitate or inhibit understanding, and that variability was apparent in learners' orientation to, and

perceived usefulness of, such visual cues. In many of the studies cited, the authors could only offer intuitive insights based on test items and responses, questionnaires, and interviews to suggest how visual content might have affected learners' comprehension. However, Ockey (2007) and Wagner (2008) specifically focused on eliciting learners' online processing of audiovisual information through verbal reports to determine the influence of the visual content on understanding in tests of academic listening ability. Ockey's study involved 6 ESL test takers who were asked to report their use of visual cues, and the impact those cues had on their comprehension during pauses inserted at essentially regular intervals in two lecture videotexts, one containing moving images, and the other still images. Five of the six test takers used hand and body gestures or facial cues in the videotext with moving images to facilitate comprehension. Few of the test takers found the still images in the lecture distracting, and all were rarely found to observe the still images in any case. There was a fairly even split between test takers who broadly found visual content either helpful, both helpful and distracting, or primarily distracting. Overall, Ockey found limited use of the still images by test takers in that version of the videotext, and that there was considerable variability in the videotext with moving images in how test takers reported utilizing the visual content, or generally considered it to be helpful or distracting.

Wagner also collected verbal reports using a pause insertion methodology. Eight ESL learners verbalized their comprehension processes at predesignated pauses as they worked through an academic dialogue and a lecture videotext, and completed the corresponding tests. Most learners reported using hand gestures in the lecture to interpret relevant parts of the videotext. In addition, several learners mentioned utilizing the body language of the speakers in the academic dialogue to help develop their interpretations of its content. Furthermore, some of the learners exploited contextual information in the academic dialogue to discern who the speakers were, and to monitor and interpret what the speakers were doing at the start of the dialogue. Similar to Ockey (2007), Wagner concluded that learners vary widely in how they attend to and exploit visual content to understand videotexts.

Although there has been comparatively less research with news videotexts than with academic lecture videotexts, it is an area that has been, and continues to be, the focus of interest for L2 researchers. For example, in several publications aimed at informing classroom practice, Meinhof (1994, 1998) describes and exemplifies the interrelations between visual and audio content in news broadcasts in terms of Overlap, Displacement, and Dichotomy. These three categories are analogous to the Direct, Indirect and Divergence categories, respectively, proposed by Walma van der Molen (2001). As in the present study, Meinhof (1994) adopts the view that by understanding how L1 users process, and are influenced by, vagaries in audiovisual content, we can come to understand their potential effects on L2 users' comprehension.

In empirical terms, Guichon and McLornan (2008) investigated aspects of multimodality (i.e., audio only, audio and visual, and the addition of L1 or L2 subtitles) in a BBC news videotext. The authors counted semantic units in learners' written summaries as a measure of what they had comprehended. In attempting to account for the differences in comprehension that were evident across the modality conditions of visual content with or without subtitles, the authors suggested that learners' comprehension may have been negatively affected at times due to a 'split-attention' effect (Chandler & Sweller, 1992), that is, the division of attention to different modes of input which increases the working memory load and reduces understanding. Importantly, and perhaps counterintuitively, a 'redundancy' effect has also been noted by Chandler and Sweller (1991), whereby the processing of simultaneous audio and visual content which is congruous has potentially negative consequences for understanding. This occurs because an increase in working memory load is associated with processing two simultaneous sources of information and attempting to establish if they are related (Sweller, 2002).

In a more extensive study of how visual content affects L2 listeners' comprehension of news videotext, Gruba (2004) investigated the ways in which learners utilized the visual content of Japanese news videotexts. Through examining the retrospective verbal reports of twelve tertiary learners of Japanese,

Gruba (2004, p. 63) identified seven aspects related to the role of visual information during news videotext comprehension:

- Listeners utilize visual elements to identify text type;
- Listeners may utilize decoded written text to form an initial macrostructure;
- Listeners may utilize visual elements to generate a number of tentative hypotheses;
- Listeners may utilize visual elements to confirm an emerging interpretation;
- The presence of a visual element may help listeners narrow an interpretation from amongst other plausible meanings;
- Visual elements may confuse or hinder interpretation;
- At times, visual elements add little to the development of a macrostructure.

Gruba (2006) also explored learners' verbal reports and semi-structured interview responses related to listening to Japanese news videotexts from a media literacy perspective, and again illustrated the influence of the visual content on listening. Regarding aspects relevant to this study, Gruba reported a case study of one learner, Abby, who was given the opportunity to replay sections of the news videotexts to create and build her understanding of content. Abby reported using visual elements to determine signposts (key visual content) and boundaries (segmentation) as a means of facilitating her search for comprehension. In addition, she became aware that aural and visual elements did not necessarily correspond. Where discrepancies existed, she attended to the audio content and ignored the visual information. When the two content sources matched, she was able to exploit this to realize greater understanding. Other learners in the study also commented in their interviews that the visual content helped reduce their anxiety, heightened motivation, and gave them a sense of connectedness with the cultural context represented on-screen.

Given this rather small body of research into news videotexts, and that only Gruba (2004, 2006) has thus far provided tangible insights into the way visual elements are processed and how they function in news videotext comprehension, there appears a need for further investigation to inform conceptual understanding and pedagogical practice, as well as generally broaden our knowledge of the influence of this key aspect of videotexts on language learners' comprehension. Thus, the research question for this study was: What is the influence of visual content on L2 listeners' comprehension of news videotexts?

## THE STUDY

### Overview

This research was part of a broader study examining the listening processes of twenty EFL learners studying at a language school in central Japan. Five BBC news videotexts were examined using Walma van der Molen's (2001) four-category coding system, and their audiovisual characteristics were accordingly quantified. A different news videotext was then utilized in each of five 90-minute lessons over five weeks. The news videotexts were edited into segments, and learners worked in pairs to complete a sequence of tasks in a pedagogical cycle for each segment (six per news videotext) at their own pace guided by a prompt sheet. The pairs did not receive any prior training in discussing their comprehension processes, nor did they receive any input from the researcher throughout the study to avoid manipulating the direction and content of their dialogue. The researcher's role was only to ensure that the pairs adhered to the task sequence and to control the playing of the news videotexts. All interaction between learners was carried out in English, reflecting the requisite use of the L2 in their regular lessons. Each pair's dialogue was audio recorded, transcribed, and acted as the unit of qualitative analysis.



## Participants

The twenty volunteers were Japanese females aged between 22 and 55. All were attending an advanced-level English language course. A comparison of course level versus IELTS band scales using the language centre's approximation table indicated that participants were at approximately IELTS band scale 7.0. All names are pseudonyms.

## Materials Preparation and Analysis

The five news videotexts used in the study were drawn from free-to-air televised BBC news broadcasts. The initial criterion for choosing the news videotexts was that they were under two minutes in length to ensure that the amount of preparation time required for editing each of the news videotexts into segments was not overly excessive. In addition, among the news videotexts selected, a range of common tradecraft features, such as interviews with members of the general public, CGTs, and CGAs, should be represented to expose learners to the typical components of this type of videotext. Furthermore, news videotexts consisting of a sequence of short segments were preferred, with each segment consisting of one or a series of images of the same scene plus accompanying audio and well-defined visual cuts between segments. This provided for ease of editing and consistency of material through the study. Each news videotext was edited into short segments according to visual scene change and shift in audio content focus, a natural discourse boundary in news videotext (see [Appendix](#)). Due to lesson time constraints, only the first six segments of each news videotext were presented in the 'classroom' phase of the study. The length of segments ranged from 6 to 22 seconds, with the average length being 14 seconds. The order of presentation and content of the five news videotexts examined in the study are summarized in Table 2.

Table 2. *Overview of the Titles, Topics and Lengths of the Five News Videotexts*

News Item Title	Topic	Length (seconds)
Term-time Holidays	School holidays in the UK	126.1
Green Grocer	Food packaging in the UK	124.5
Elderly Abuse	Aged-care in the UK	129.9
Job Losses	Unemployment in the UK	118.3
Basra Deaths	UK forces in Iraq	130.5

To explore the nature of the information presented in these five news videotexts, verbal and visual content in each videotext for all segments was analyzed according to the four categories suggested by Walma van der Molen (2001)—Direct, Indirect, Divergent, and Talking head. The coding method employed was similar to Walma van der Molen's. A coding form was prepared in which the verbal script was given alongside images of associated shots and a brief written statement describing the shots. During the analysis, one of the four coding categories was selected and noted on the coding form for the given audiovisual information. However, Walma van der Molen used shot duration to calculate the time for each category. Instead, in this study, the time for a category was calculated based on the duration of the utterances accompanying associated shots. Pauses before and after utterances were excluded, as only visual content was being presented. Measurements were made using Praat Version 4.3.22 (<http://www.fon.hum.uva.nl/praat/>). A colleague acted as a second coder for coding agreement checks. An inter-coder reliability analysis using the Kappa statistic was performed with SPSS Version 18.0 (<http://www.spss.com/>). Inter-coder reliability was .79. Differences in coding were then resolved by discussion to enable the analysis of the prevalence of each of the four categories.

## Elicitation and Analysis of Dialogue

To provide the framework for eliciting learners' dialogue for subsequent analysis based on the first six



segments of each news videotext, a pedagogical cycle proposed by Vandergrift (2007) was used (see Cross, in press, for details). Pairs watched a segment on a TV set, and then made notes after the segment finished. Next, learners shared their understanding of the segment, discussed how they had tried to understand the content, and considered ways to understand more of the segment. Specific written prompts were provided to elicit learners' responses, such as, "What strategies did you use to try to understand the segment?" The learners discussed their comprehension processes at designated pauses inserted in the news videotext, akin to the manner in which Ockey (2007) and Wagner (2008) collected verbal protocols. The same segment was then replayed, learners added to their notes, shared their understanding, and reported on how they had tried to understand the segment. Following this, learners worked together to produce a written summary of main ideas they had jointly comprehended. On average, they spent approximately fifteen minutes working on each segment. On finishing a segment's summary, the learners signaled to the researcher to play the next segment.

The qualitative analysis by the author of each pair's 'dialogic recalls' (Cross, 2011) using QSR NVivo Version 8 (<http://www.qsrinternational.com/>) aimed to establish the influence of the visual content on their comprehension of the given news videotexts. Excerpts in which a learner's report referred to the visual content were firstly identified and grouped for each news videotext. These excerpts of dialogue were then individually cross-referenced to the coding form (see previous section) to establish the category of the relationship (i.e., Talking head, Direct, Indirect, and Divergent) of the audiovisual content that had been the focus of learners' dialogue. Excerpts of dialogue related to the visual content in the news videotexts which were more general in nature and could not be linked to any of Walma van der Molen's four coding categories, were collated and given provisional labels for each pair. Excerpts with related labels were then matched across the ten pairs, and the categories iteratively consolidated. Isolated excerpts which could not be cross-matched were excluded from further consideration. A colleague again acted as a second coder, and was asked to use the categories established (i.e., positive or negative effect, inferencing, and predicting) to code the excerpts which did not relate to the coding form content. An inter-coder reliability analysis using the Kappa statistic was conducted with SPSS. Inter-coder reliability was .83. Coding differences were subsequently resolved by discussion.

## FINDINGS AND DISCUSSION

The initial concern of this study was to draw on L1 research to establish the nature of the audiovisual content in the BBC news videotexts utilized. Using Walma van der Molen's (2001) audiovisual coding system, all of the segments in each of the five news videotexts used in this study were analyzed and the amount of time in seconds and as a percentage of the total time for each of the four categories was quantified. [Table 3](#) shows the findings of the analysis for each of the news videotexts.

[Table 3](#) highlights that the Talking head category was the most common, with each news videotext containing two or three of such segments. Overall, the majority of the audio and visual content for each of the five news videotexts was classified as Indirect or Divergent. This supports the generally held view that redundancy between audio and visual modes in news videotexts is rare (Meinhof, 1998; Walma van der Molen, 2001). Two of the news videotexts contained a CGT segment (Elderly Abuse and Job Losses) and one a CGA segment (Basra Deaths), and these segments exhibited a notable level of audiovisual correspondence. Hence the higher percentages for the Direct category compared to the other two news videotexts (i.e., Term-time Holidays and Green Grocer). The implication for the L2 listeners in this study was that the prevalence of at least partial audiovisual discrepancy in the five BBC news videotexts clearly had the potential to create comprehension difficulties, much as it had affected uptake in the L1 study by Brosius, et al. (1996) discussed above.

To examine the manner of influence of the audiovisual content on learners' comprehension, excerpts of dialogue from each pair were cross-referenced to the relevant shots they were referring to in their dialogue, and the coding of those shots as Talking head, Direct, Indirect or Divergent. The influence of

the visual content on learners' comprehension is now discussed in terms of each of these four audiovisual correspondence categories.

Table 3. *The Types of Visual Content and Their Distribution in Seconds and as a Percentage for Each of the Five News Videotexts*

News Item Title	Types of visual content	Distribution of visual content	
		(seconds)	(%)
Term-time Holidays	Talking head	41.7	37
	Direct	2.9	3
	Indirect	30.0	26
	Divergent	38.8	34
	Total	113.4	100
Green Grocer	Talking head	39.0	35
	Direct	0.0	0
	Indirect	33.2	30
	Divergent	39.6	35
	Total	111.8	100
Elderly Abuse	Talking head	44.9	41
	Direct	15.7	14
	Indirect	20.8	20
	Divergent	27.0	25
	Total	108.4	100
Job Losses	Talking head	35.9	33
	Direct	19.9	18
	Indirect	17.3	16
	Divergent	35.3	32
	Total	108.4	100
Basra Deaths	Talking head	24.6	24
	Direct	20.6	20
	Indirect	36.0	34
	Divergent	23.4	22
	Total	104.6	100

*Note: Pauses in audio content, which meant only visual content was presented, were excluded from time calculations in each category.*

### Talking Head

This category refers to close-up shots of the head and upper body of newscasters, reporters, and interviewees which did not contain background scenes of semantic significance. In this study, each of the news videotexts began with a Talking head segment in which the newscaster can be seen introducing the news story. Four of these segments also included a caption identifying the title and, therefore, the theme

of the news videotext (see [Appendix](#), Segment 1). In addition, three of the news videotexts included Talking head segments in which an interviewee offered their views. Each interviewee was identified with a caption (see [Appendix](#), Segment 3). None of the learners' reports contained comments about the Talking head shots of the newscasters. Thirteen learners (in ten excerpts) mentioned the visual content in segments with Talking head shots of interviewees. Three learners reported just focusing on the audio content as the visual content in the Talking head segment was not felt to be semantically informative. For instance, Nao reported that the Talking head shot merely showed the woman talking (see [Appendix](#), Segment 3), and so she had concentrated on what the interviewee was saying.

- Nao: uhm about for the visual points  
 Midori: mm  
 Nao: a woman is just talking about  
 Midori: mm mm  
 Nao: just talking, so I concentrated on the words  
 Midori: mm  
 Nao: I I I can catch

However, two learners reported using the captions identifying an interviewee to orientate themselves to who the actual speaker was. Interestingly, three learners also reported not noticing the captions, despite their appearance on screen for most of the duration of the given segments. An excerpt from Azusa and Yoko's dialogue with respect to the female interviewee (see [Appendix](#), Segment 3) illustrates both of these aspects, with Yoko mentioning she used the caption, whereas Azusa reported not seeing it.

- Yoko: ah I first of all who is speaking  
 Azusa: uhuh  
 Yoko: the head teacher  
 Azusa: eh eh how did you know she's the head teacher  
 Yoko: the subtitles subtitles  
 Azusa: [ah you saw the subtitles I didn't see that point  
 Yoko: and ah this is the head teacher

In summary, several learners' reports reflected that they felt visual content in Talking head shots provided little of semantic value to facilitate their understanding, and thus they tended to direct their attention to the contiguous aural content. The captions identifying interviewees did help to orientate a few learners to the name/position of the speaker, but this feature could also go unnoticed.

### Direct Category





This category describes instances in which there was a high degree of semantic equivalence between aural and visual information. Apart from excerpts linked to the CGT and CGA segments, there was one excerpt each from the dialogue of four pairs in which learners discussed the influence of audiovisual content classified in this category. All of the learners mentioned that the visual content had supported their understanding of audio content in relation to a scene from *Basra Deaths* in which parts of a vehicle destroyed in an explosion are seen on the road (see [Table 1](#)). For example, Jun mentions that seeing the wreckage on the road had facilitated her understanding of this part of the segment.

- Jun: uhm if ah when I saw the pic- erh the image of erh the wreckage parts are strew- strewn around across the road
- Kaori: [uhuh [ah yes yes
- Jun: that was very helpful to understand the the what's happening
- Kaori: [mm yes mm
- Jun: at that time
- Kaori: yes on the road

This part of the segment was notably short (1.9 seconds), yet the audiovisual correspondence appeared to be particularly apparent to several of the learners, and drew comment. It is unclear as to why no other excerpts were related to audiovisual content in the Direct category (other than for the CGT and CGA segments). It may be that the brief duration of each example of such content in general (the average time of audiovisual content in the Direct category was 2.6 seconds), or that just under half of the examples were only a partial component of a proposition (e.g., see the two examples of the Direct category in [Table 1](#)), made it difficult for learners to recognize and exploit the semantic overlap in the audio and visual content. Alternatively, audiovisual redundancy may not have been recalled as associated content was unconsciously processed, or because it was one small part of the complex process of comprehending a segment's propositional content (typically each segment contained three propositions).

The CGT and CGA segments contained audiovisual content which exhibited redundancy. The CGT segment in Elderly Abuse consisted of a sequence of numbers and on-screen text, and around half of this segment's content exhibited semantic overlap. Table 4 shows the content which was categorized as Direct.

Table 4. *Examples of the Direct Category in the CGT from Elderly Abuse*

Verbal content	Visual content	
In more than two hundred cases the person was abused in their own home.		
There have been just five prosecutions.		

Seven pairs commented on the effect on their comprehension of the CGT visual content in ten excerpts. This was primarily regarding facilitating understanding of numerical details which is typically difficult for L2 listeners. For example, Yoko reported using the number graphic, recognizing it was linked to the audio content, and thus being able to comprehend the information presented.

- Yoko: I tried to follow the numbers appearing on the screen
- Azusa: [mm uhuh
- Yoko: and the sound is connected with that number
- Azusa: uhuh

Yoko: so I could understand what this number is what

The CGA in Basra Deaths, shown in Table 5, also contained audiovisual content which was redundant. This segment primarily portrayed an explosive attack on a vehicle, with the voiceover describing how such an attack proceeds.

Table 5. *Examples of the Direct Category in the CGA from Basra Deaths*

Verbal content	Visual content	
But shaped bombs are designed to focus the force of the explosive into a small area, forcing a hard projectile through the light armour of a Snatch Land Rover.		
		

Excerpts of dialogue from seven pairs illustrated that this CGA appeared to have either a positive or negative influence on learners' comprehension. Of eighteen related excerpts, ten were positive. For instance, both Emi and Kana reported that their understanding of the military technology shown in the segment had been facilitated by the animation.

Emi: yes in my case I I watched the illustration that something weapons attacked to the land rover mm:: so I think it helps me to understand the weapons how weapons how sophisticated the weapon mm

Kana: mm yeah erh well in my case I uhm thanks for the clear illustration illustrations I thought I could understand the basic concepts of the weapons

However, Masako and Satsuki were among the learners who reported the CGA had inhibited or impaired their comprehension irrespective of audiovisual correspondence because of the nature and amount of information it contained. Satsuki reported being absorbed in the visual content and forgetting to attend to the audio content, while Masako stated that her attempts to exploit more of the visual content had led to increased confusion.

Satsuki: I was I was I was so so attracted by the scene

Masako: mm

Satsuki: the truck land rover and the explanation and illu- and illus- illustration

Masako: mm mm mm

Satsuki: and and so I forgot to listen to erh what announcer said

Masako: [oh:: mm





- mm mm mm
- Satsuki: so
- Masako: ok so I I tried to get more information
- Satsuki: mm
- Masako: from screen
- Satsuki: mm
- Masako: but ah:: mm:: it made me more confused confused

Overall, there was a fairly even split between the number of learners who reported that the graphics in the CGA had been beneficial to their comprehension or had impaired it. The difficulty for learners seemed to be in concurrently coordinating their attention, decoding and integration of the on-screen animation and the details presented aurally. This was a procedure which possibly overwhelmed their cognitive resources. Therefore, it seems that despite redundancy between audio and visual content, the sheer volume of information from different sources (i.e., written text, audio, animated visual scenes) in CGAs, which is designed to assist L1 users' understanding of complicated events or processes in news videotexts, could possibly confuse some L2 listeners and make it difficult for them to build connections between audio and visual sources of information. Alternatively, learners' cognitive resources could have been overloaded as they tried to establish that correspondence existed between the multiple sources (i.e., a 'redundancy' effect, see Chandler & Sweller, 1991; Sweller, 2002). This did not appear to be as problematic with the CGT as the visual content consisted only of written text accompanied by redundant audio information, so it may be that the 'moving picture' aspect of CGAs adds an extra element of complexity for learners.

### Indirect Category

This category refers to audio and visual content which has partial semantic redundancy. One example of this type of audiovisual correspondence was when a reporter was seen using hand gestures and simultaneously referring to on-screen items or locations. This use of gestures by a reporter occurred in two of the five news videotexts. For instance, in a segment from the Green Grocer news videotext shown in Table 6, the reporter is seen holding and gesturing towards some packaging and a plastic bag as he is talking about them.

Table 6. *Examples of the Hand Gestures from Green Grocer*

Verbal content	Visual content	
Sainsbury's argue that the natural products in this packaging will break down very quickly in compost...		
...whereas the degradable plastic used by some rivals is still oil-based and will take a couple of years to break down completely.		

Four pairs discussed the visual information in this segment. For example, Tomoko reported on the way in which the visual content influenced her understanding. She stated how the reporter had explicitly drawn

attention to objects using his hands, and how this had helped her to recognize that he was making a comparison between objects.

Tomoko: when they did some comparison between sainsbury's and other retailers rivals like when he talked about the sainsbury's products he used I I think he

Nami: mm:: uhuh

Tomoko: drew up our attention to the sainsbury's products and when he talked about plastic bags or biodegradable bags by other retailers he hold the bags visually we could notice that he was comparising

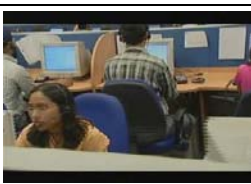
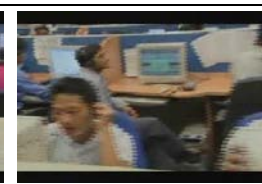
Nami: mm::

Tomoko: huh comparising comparing sorry comparing something with something

This representative example illustrates that the semantic overlap achieved through the use of hand gestures for comparing and contrasting by the reporter helped to orientate some of the learners to, and facilitate their understanding of, the aural content<sup>2</sup>. This is in line with Wagner's (2008) findings that hand gestures can help learners to interpret information in videotexts, and supports the perceptions of the learners in Coniam's (2001), Ockey's (2007) and Sueyoshi and Hardison's (2005) studies regarding the usefulness of a speaker's gestures in aiding listening comprehension.

Visual content in the Indirect category in the form of standard scenes (i.e., visual content which has a thematic correspondence with the audio content) informed learners' contextual/thematic orientation. All ten pairs commented on this aspect in relation to various segments of each videotext, and there were thirty three associated excerpts in their dialogues. For instance, in Job Losses, the visual content shows employees at work in a call-centre in India, as shown in Table 7, and the audio information is about office jobs being shifted to India.

Table 7. *Example of the Standard Scene from Job Losses*

Verbal content	Visual content	
Indian workers able to do the same office jobs more cheaply.		

Manami and Keiko discussed this visual content, and their dialogue illustrates that it had enabled Manami to achieve situational orientation. She reported she was able to notice the disparity in the visual scene and this had helped her recognize the related context of the information being presented in the segment.

Manami: the first thing I noticed is that the the visual was very parallel to to the one we saw in the segment four

Keiko: parallel



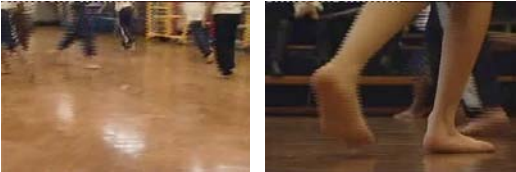

- Manami: it was similar but different to the one taken in Britain
- Keiko: mm
- Manami: it was the similar office
- Keiko: mm mm mm mm
- Manami: but something was different
- Keiko: mm
- Manami: people were different the partition and the configuration  
were was different so in a way it helped me that to to to  
notice that this is the situation in India or the uhm the  
exported situation

In summary, excerpts related to the reporter's use of hand gestures indicated that this aspect helped to orientate several learners to items or locations being depicted. In addition, the presence of standard scenes appeared to have a positive influence on some learners' comprehension through activating and informing contextual/thematic orientation and helping them to refine their interpretations of the given news videotexts as they formed and developed a macrostructure representation (Gruba, 2004).

### Divergent Category

With respect to this category, which refers to audio and visual content which is unrelated or possibly contradictory, there were only four related excerpts evident in the pairs' dialogues. Two pairs commented on audiovisual divergence when it was patently apparent. A segment in Term-time Holidays, shown in Table 8, contains visual content of children running around a gym mostly showing their legs, while the audio information is an explanation of a court case.

Table 8. *Examples of the Divergent Category from Term-time Holidays*

Verbal content	Visual content
The issue has been brought to the fore again because of a court case involving a mother from Kent.	
She was prosecuted after taking her children on holiday twice without their schools permission. After a legal battle, the High Court ruled that she'd broken the law.	

One example illustrates Azusa mentioning to Yoko how the divergent visual and audio information affected her ability to concentrate and listen, and how she felt the visual content of the children running around had hindered her concentration and comprehension.

- Azusa: I think I watched very concentratedly con- co- concentrate  
on the tv so

- Yoko: [mm  
 Azusa: that's why I couldn't catch a lot  
 Yoko: mm::  
 Azusa: because I wh- while I was listening I always thought what is  
 what are those why are they running running so that bothered  
 my concentration and listening

This representative excerpt illustrates that when the audio and visual content is particularly incongruous and also, in this case, peculiar in terms of the camera technique, it becomes apparent to some learners and can create confusion. A possible reason for why few excerpts of dialogue related to content classified as Divergent is that, although other segments contained disparate audio and visual content, the storyline of the audio and the associated visual images (excluding Talking head shots) were related to previous segments in the news videotext, and it is possible that learners were able to orientate themselves to the continuing thread of the storyline as their tentative macrostructure of the news videotext evolved (Gruba, 2004). This was not the case for the segment content in [Table 8](#), and may have been why learners felt it was problematic.

### Other Influences

In the qualitative analysis of each pair's dialogue, a number of other general influences related to the visual content emerged across a number of pairs. Firstly, in twenty three excerpts, learners in all pairs mentioned in broad terms that they found that visual content had facilitated their comprehension at some stage, as in this example:

- Satsuki: we:: this time uhm the visual points helped us very much I think  
 Masako: [mm [mm:: ah yes  
 Satsuki: mm:: and erh I think uhm we we cou- we have got a lot of information  
 Masako: mm  
 Satsuki: erh from the visual points  
 Masako: [mm mm mm yes yes

However, this needs to be qualified, as there were eighteen excerpts in the dialogue of seven of the pairs which illustrated that when they attended to the visual content in a segment, it impaired their ability to attend to the accompanying audio content. For instance, Hiromi stated that she recognized that her attention to the visual information had inhibited processing of the concurrent audio input.

- Hiromi: mm:: mm:: I tried concentrate only on the screen so  
 actually sound did not enter  
 Naoko: really  
 Hiromi: [my brain  
 Naoko: ah::

Another interesting aspect was that the initial scene in a segment (i.e., the post-cut shot) was used by learners to generate expectations about possible audio content. There were fourteen excerpts related to this in the dialogue of five pairs. For example, Manami reported that the initial image of a child in a segment from Term-time Holidays (see [Appendix](#), Segment 2) had created an expectation regarding the context, which she felt had assisted her comprehension.

Manami: ah:: but you know when I first see I first saw the head of  
of a child I I immediately

Keiko: [mm

Manami: knew they were going to talk about the classroom and it  
helped me

Keiko: [mm:: ah::

A further influence of the visual content was that it aided inferencing by learners of a segment's information. There were thirteen related excerpts among six pairs. One example shows Emi reporting that the visual scene had been the stimulus for guessing the content. She mentioned concentrating on the visual content and seeing parts of a vehicle on the road in Basra Deaths (see [Table 1](#)), and using this information to conclude that the vehicle had exploded.

Emi: erh I watched on the screen carefully and yes I I saw a car

Kana: [uhuh mm

Emi: and some metal things such as coil

Kana: [mm mm

Emi: and the metal plate on on the on the place so I I guessed the cars

Kana: [mm::

Emi: exploded exploded and mm::

In summary, learners considered that visual content could both promote and impede their understanding of the accompanying verbal material. Other studies have also found in broad terms that visual content can be both helpful and distracting. Regarding the latter, as with a number of the participants in Coniam's (2001) and Ockey's (2007) studies, the visual content seemed to exclusively absorb learners' attention at times, causing them to fail to allocate resources to processing the simultaneous audio information. In addition, using the initial visual content of a segment helped learners to predict or create expectations about the possible information presented in that segment. Although this could be a risky strategy, just over half of post-cut shots were good indicators of the focus of segment content. A further strategy learners felt had facilitated their comprehension was inferencing based on the visual content (see also Gruba, 2004). The visual content possibly provided a tentative frame of reference which learners used to organize the parts of the audio content they could comprehend and create coherent propositions. Of course, this does not imply that their inferences were necessarily always correct, especially as there was a high proportion of content in the news videotexts that lacked audiovisual redundancy.

### Learner Variability

Ockey (2007), Sueyoshi and Hardison (2005), and Wagner (2008) have commented that the influence of the visual content on comprehension of videotexts is notably variable for each learner, and the findings of this study also illustrate that this is so. While L2 listening comprehension is primarily an idiosyncratic process and, as such, one would expect differences to be evident among learners regarding their frequency and degree of use of visual content in videotexts (as reflected in their verbal reports), it is informative to account for how such variability possibly arises. Wagner (2008) suggests that one reason for the variability is because visual content tends to be automatically processed, so learners are not conscious of doing so. Hence, it is not available for reporting by the given learner. However, it is debatable that the socioculturally-bound visual content in videotexts requires little conscious effort on the part of L2 learners to extract the semantic notions being conveyed. Rather, the analysis of each pair's

dialogue in this study revealed that some variability can be more plausibly explained from a dual coding theory perspective, which advocates distinct verbal and nonverbal systems (Paivio, 2007). Given that the multimodality of news videotexts places excessive demands on an individual's limited short-term memory capacity (Lang, 1995), a number of learners appeared to employ a coping mechanism in which they intentionally directed their attention to either the visual content or the audio content, with the incumbent loss of information presented in the non-attended content source. Moreover, eight learners reported deliberately switching their attention across the two listenings to each segment, primarily attending to the visual content in the first listening to a segment, and focusing on the audio content in the second listening. As such, these learners were likely to comment on the visual content following the first listening only, particularly when they did not find the visual content initially useful. For example, Masako reported that the visual information had not aided her understanding during the first listening so she had decided to attend to the audio content in the second opportunity to listen.

- Masako: mm:: on the screen there is no there's no tips I mean  
 Satsuki: yeah  
 Masako: hints  
 Satsuki: yeah mm  
 Masako: so it's quite difficult to  
 Satsuki: mm  
 Masako: mm to guess from the visual in the part  
 Satsuki: [yeah mm yeah  
 Masako: so next time I'm going to concentrate on erh the the listening  
 Satsuki: mm::

Similarly, in the following excerpt, whereas Manami mentioned using visual information in the second listening to a segment, Keiko reported that she had consciously not attended to the visual content the second time she listened to the segment and had focused on the audio content.

- Manami: this time I tried to use the visual  
 Keiko: mm mm mm and I yeah I this time I ignored the visual  
 Manami: [right yeah  
 Keiko: and concentrate to hear

Gruba (2004) also noted the tendency of learners to primarily attend to the visual content in the first listening to formulate an initial impression, and then develop a more complete understanding by attending to the aural content as they listened again. Therefore, it appears that conscious attention to either, but not simultaneously to both, the audio or visual content is a way learners attempt to overcome processing issues they encounter, such as when a 'split attention' or 'redundancy' effect overwhelms their short-term memory resources.

In relation to learners focusing their attention on either source of content, there was evidence that while all learners were observed to 'look' at the screen when listening, this was, at times, possibly a 'blank stare'—an unfocused look that does not involve the processing of what is seen (Garland-Thomson, 2009). The visual information in the news videotext was not necessarily being utilized, and learners concentrated on processing the audio content alone. For example, the following excerpt illustrates that Naoko adopted this type of behavior.

Naoko: so some sometimes I I point my eyes on the screen but not exactly focus on

Hiroshi: ah:: blankly you look at ok

Naoko: yeah so next

These findings have implications for studies in which researchers have measured the time learners spent observing the visual content (e.g., Ockey, 2007; Wagner, 2007, 2010a). Although learners are seen to be orienting to the screen, this study shows that it does not necessarily mean they are attending to and exploiting (consciously or unconsciously) the visual elements displayed.

Overall, then, it appears that the visual content in news videotexts, irrespective of the degree of audiovisual correspondence, creates a further significant strain on learners' limited cognitive resources. Learners may try to deal with this issue through directing their attention at different times to information from only one source in preference to the other. This may depend on which source the individual learner feels can best be effectively exploited to interpret and ascertain meaning in the news videotext, and seems to be one important reason for why variability in the use of visual content exists across learners<sup>3</sup>.

### IMPLICATIONS FOR PEDAGOGY

Despite a number of limitations of this study, including the participants being a rather homogenous group, the news videotext segmentation possibly distorting normal discourse processing, and potential issues with using dialogic recalls as verbal reports (see Cross, 2011), several implications for L2 listening pedagogy arise from the findings. Firstly, it was evident that not all learners recognized congruence and discrepancies between the aural and visual elements as they strove for understanding. This suggests that such aspects need to be made explicit to learners if they are to better deal with the audiovisual vagaries of news videotexts. One technique for achieving this is to present learners with a range of segments and ask them to compare the transcript of the aural content with the visuals they see, and determine and discuss the extent of audiovisual correspondence. Another approach is to have learners predict the kind of visual content they think corresponds to the transcript of the audio content of a news videotext, and then ask them watch the videotext and reflect on the degree of audiovisual correspondence that was evident. In particular, it seems apposite to raise learners' awareness of the utility of hand gestures used by reporters, and the potentially facilitative effect of numbers and/or captions presented in Talking head, CGT and CGA segments. Furthermore, several studies have shown the facilitative nature of speakers' lip movements and facial expressions for understanding (Ockey, 2007; Sueyoshi & Hardison, 2005). Talking head segments are a common element of news videotexts, and it would be useful to draw learners' attention to such features with respect to newscasters and interviewees.

In addition, from a media literacy standpoint, Gruba (2006) points to the importance of learners being able to identify segment boundaries using visual elements. This skill helps listeners to keep pace with shifts in content focus as the news videotext progresses. For this study, boundaries were predetermined to enable separate segments of the news videotexts to be presented one-by-one. However, in a classroom context, learners could initially be introduced to the notion of segmentation through visuals in news videotexts and how it operates. A complete news videotext could then be presented (with or without sound) and learners asked to discuss and justify the number of segments they feel it contains. This could be facilitated by asking them to mark segment boundaries on a transcript of the news videotext. Drawing learners' attention to the regularity of the generic macrostructure of news videotexts (Meinhof, 1998) is also important for developing media literacy. In the BBC news videotexts used in this study, we first see the anchor in the studio, and this is followed by a sequence of short multimodal segments. Viewpoints presented are supported through interviews with 'stakeholders' (i.e., members of the general public, politicians, victims), and the final segment often shows the correspondent at the scene or contains the correspondent's voiceover indicating current consequences and future directions. Of course, it is important to note that the macrostructure and content of BBC news videotexts are culture-bound. In

contrast, Japan's NHK news, for example, has a different macrostructure and content (see Botting, 2003). In addition to raising awareness of the macrostructure, learners' knowledge of the defining features of news videotexts according to particular themes (e.g., politics, war, crime) may be enhanced by using worksheets to guide and maximize their listening experience (see Lynch, 2009). Furthermore, as Gruba (2005) suggests, learners can predict the meaning of the visual content in a news videotext and compare their ideas. Through doing so, they can become aware that visual content may have 'polysemic' interpretations (i.e., an array of diverse meanings, Gruba, 2005).

## CONCLUSION

This study identified and examined the various audiovisual characteristics of (BBC) news videotexts using a four-category system and coding method developed by Walma van der Molen (2001). It was evident that audiovisual correspondence in the news videotexts was non-equivalent to varying degrees. Subsequent analysis focused on learners' dialogue to explore the effect the four different categories of shot types had on learners' listening comprehension. Talking head visual content seemed to have little influence on comprehension, though captions did help with speaker identification. The effect of the visual content classified as Direct was typically facilitative of comprehension, but the multimodality of contiguous information in CGAs could be detrimental to understanding. Indirect audiovisual correspondence, as reflected in the hand gestures of the reporter and in standard scenes, influenced comprehension positively, whereas Divergent audio and visual content seemed particularly problematic when it was notably incongruous with the evolving news videotext storyline.

In addition, the analysis revealed other influences of the visual content on comprehension such as its role in facilitating comprehension; inhibiting of attention to, and processing of, audio content; and triggering of learners' expectations and inferencing of content. Dual coding theory provided a useful perspective for explaining possible reasons for why there is notable variability among learners in the degree to which they report exploiting the visual content in news videotexts, and it is hoped the implications for L2 listening pedagogy presented offer a way forward for practitioners using news videotexts (or other types of videotexts) in their listening lessons.




---

## NOTES

1. Listening comprehension is defined here as "an active process in which listeners select and interpret information which comes from auditory *and visual* [this author's italics] clues" (Rubin, 1995, p. 7).
2. The audiovisual correspondence was coded as Indirect as the visual information presented includes the supermarket interior, the shopping aisle, and items in a trolley.
3. Other potential reason for learner variability in reports of their use of visual content include the tendency for this information to evoke polysemic interpretations (Gruba, 2005), and the disparate visual literacy, spatial ability, and background knowledge of learners (Chun & Plass, 1997).



**APPENDIX. Example of a segmented news videotext**

Segment verbal content	Segment visual content
<p style="text-align: center;">Segment 1</p> <p>Now, if you think you could save a tidy sum by taking your kids on holiday in term time, you could be in for a nasty surprise. In a test case, the High Court has backed the law which says it's schools who decide if these trips are OK. So, what exactly are parents allowed to do? Judith Morris has been finding out.</p>	
<p style="text-align: center;">Segment 2</p> <p>These children at school in Manchester are all present and correct, but that's not the case everywhere. Most teachers marking the register have had the experience of pupils taking time off to go on holiday. It can be a tug of war between parents and schools.</p>	
<p style="text-align: center;">Segment 3</p> <p>It's escalating in the number of families that are actually taking children out of school. Parents now ...erh... expect to take probably more than one holiday a year. I do have a sympathy with parents because the guidelines are not clear. And it's left too much onto head teachers.</p>	

**ACKNOWLEDGMENT**

I wish to acknowledge Major Matthew Bacon, who is mentioned in the Basra Deaths news videotext used in this study, and who lost his life on 11th September, 2005 while serving in Iraq.



## ABOUT THE AUTHOR

Jeremy Cross is an Assistant Professor of Applied Linguistics at the National Institute of Education, Nanyang Technological University, Singapore. His primary research interest is L2 listening. He mainly teaches postgraduate courses on ELT methodology for listening and speaking.

E-mail: [jertz7@hotmail.com](mailto:jertz7@hotmail.com)

---

## REFERENCES

- Botting, G. (2003, February 18). Japan's TV news in a world of its own. *The Japan Times Online*. Retrieved from <http://search.japantimes.co.jp/cgi-bin/fl20030218zg.html>
- Brinton, D., & Gaskill, W. (1978) Using news broadcasts in the ESL/EFL classroom. *TESOL Quarterly*, 12(4), 403–414.
- Brosius, H-B., Donsbach, W., & Birk, M. (1996). How do text-picture relations affect the informational effectiveness of television newscasts? *Journal of Broadcasting and Electronic Media*, 40, 180–195.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233–246.
- Chun, D., & Plass, J. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, 1(1), 60–81. Retrieved from [http://llt.msu.edu/vol1num1/chun\\_plass/default.html](http://llt.msu.edu/vol1num1/chun_plass/default.html)
- Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System*, 29, 1–14.
- Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research*, 13(2), 151–176.
- Cross, J. (2011). Utilizing dialogic recalls to determine L2 listeners' strategy use. *Innovation in Language Learning and Teaching*, 5(1), 81–100.
- Cross, J. (in press). Metacognitive instruction for helping less-skilled listeners. *ELT Journal*.
- Fox, J., Lang, A., Chung, Y., Lee, S., Schwartz, N., & Potter, D. (2003). Picture this: Effects of graphics on the processing of television news. *Journal of Broadcasting and Electronic Media*, 48(4), 646–674.
- Garland-Thomson, R. (2009). *Staring: How we look*. Oxford: Oxford University Press.
- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133–167.
- Graddol, D. (1994). The visual accomplishment of factuality. In D. Graddol & O. Boyd-Barrett (Eds.), *Media texts: Authors and readers* (pp. 136–59). Clevedon, UK: Multilingual Matters.
- Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning*, 17(1), 51–82.
- Gruba, P. (2005). *Developing media literacy in the L2 classroom*. Sydney: Macquarie University,

National Centre for English Teaching and Research.

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, 10(2), 77–92. Retrieved from <http://llt.msu.edu/vol10num2/pdf/gruba.pdf>

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85–93.

Gunter, B. (1987). *Poor reception: Misunderstanding and forgetting broadcast news*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Harmer, J. (2001). *The practice of English language teaching*. Essex, UK: Longman.

Lang, A. (1995). Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research*, 22, 86–115.

Lynch, T. (2009). *Teaching second language listening*. Oxford: Oxford University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83, 484–490.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86, 389–401.

Meinhof, U. (1994). Double talk in news broadcasts: A cross-cultural comparison of pictures and texts in television news. In D. Graddol & O. Boyd-Barrett (Eds.), *Media texts: Authors and readers* (pp. 212–223). Clevedon: Multilingual Matters.

Meinhof, U. (1998). *Language learning in the age of satellite television*. Oxford: Oxford University Press.

Ockey, G. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517–537.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical interpretation*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Reese, S. (1984). Visual-verbal redundancy effects on TV news learning. *Journal of Broadcasting*, 28, 79–87.

Rivens-Monpean, A., & Guichon, N. (2009). Assessing the use of aids for a computer-mediated task: Taking notes while listening. *JALT CALL Journal*, 5(2), 45–60.

Rubin, J. (1995). An overview to *A guide for the teaching of second language listening*. In D. Mendelsohn & J. Rubin (Eds.), *A Guide for the teaching of second language listening* (pp. 7–11). San Diego, CA: Dominic Press.

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Sherman, J. (2003). *Using authentic video in the language classroom*. Cambridge: Cambridge University Press.

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55, 661–699.

Sweller, J. (2002). Visualisation and instructional design. Workshop paper for the IWM Knowledge Media Research Center. Retrieved from <http://www.iwm-kmrc.de/workshops/visualization/sweller.pdf>

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191–210.

Wagner, E. (2007). Are they watching? An investigation of test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, 11(1), 67–86. Retrieved from <http://llt.msu.edu/vol11num1/wagner/default.html>

Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, 5, 218–243.

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System*, 38(2), 280–291.

Wagner, E. (2010b). How does the use of video texts affect ESL listening test-taker performance? *Language Testing*, 27(4), 493–510.

Walma van der Molen, J. (2001). Assessing text-picture correspondence in television news: The development of a new coding scheme. *Journal of Broadcasting and Electronic Media*, 45(3), 483–498.

Walma van der Molen, J., & Van der Voort, T. (2000). The impact of television, print, and audio on children's recall of the news: A study of three alternative explanations for the dual-coding hypothesis. *Human Communication Research*, 26(1), 3–26.