AUDIBLE HYPERLINKS IN SYNTHETIC SPEECH: EFFECTS OF SPEECH AND NON-SPEECH CUES ON HYPERLINK PERCEPTION & SENTENCE COMPREHENSION

Dan Hamer-Hodges, Simon Li and Paul Cairns

University College London Interaction Centre,
University College London
London,
UK.
dan@geekboy.co.uk, simon.li@ucl.ac.uk, p.cairns@ucl.ac.uk

ABSTRACT

This paper describes two empirical experiments investigating the perception of embedded audible hyperlinks, designed using speech and non-speech cues, and their effect on the comprehension of synthetic speech. Results from the first experiment showed high accuracy levels of hyperlink perception and differences in comprehension performance between sentences with hyperlinks and sentences without hyperlinks. Results from the second experiment also showed high accuracy levels of hyperlink perception as well as differences in comprehension performance between two hyperlink designs using different configurations of speech and non-speech cues.

The results demonstrate that speech and non-speech cues may be effective in the design of audible hyperlinks however their presence within synthetic sentences may reduce overall comprehensibility. Results also demonstrate that different configurations of speech and non-speech cues used to represent audible hyperlinks effect comprehension processes.

1. INTRODUCTION

Hypertext is multi-dimensional, structured information that has its roots in the visual medium. Web pages combine text information with meaningful semantic mark-up of the text (King et al., 2004) which, when viewed in a graphical web browser, explicitly represents both the textual and structural contents of a document (James, 1996). As the auditory equivalent of visual hypertext, audible hypertext raises challenge for designers who must find ways of communicating the structural information about content without the benefit of a visual display (James, 1996, Wynblatt et al., 1997; Morley et al., 1998; Goose et al., 2000; Goose et al., 2002; King et al., 2004). Users of audible hypertext must be able to correctly *perceive* and *comprehend* these two dimensions of the user interface without the benefit of visual redundancy. The implication is that the usability of audible hypertext systems depends, in part, on the ability of interaction designers to create an audible user interface that listeners can perceive and comprehend with relative ease.

Embedded hyperlinks present a special challenge in terms of perception and comprehension because they must be sufficiently intelligible to be perceived within a passage of speech and sufficiently unobtrusive to ensure the listener's comprehension of the surrounding material. This challenge is made more difficult by the fact that audible hypertext speech output uses synthetic speech, which has been shown to be less intelligible and less comprehensible than natural speech (Luce, Feustel & Pisoni, 1983; Pisoni, Manous & Dedina, 1987; Ralston, Pisoni, Lively, Greene & Mullennix, 1995; see also Ralston, Pisoni & Muliennix, 1995 for a review).

Audible hypertext content is becoming increasingly available in commercial desktop applications and over-the-telephone systems. Voice browsers (IBM, 2004) and screen-readers (JAWS, 2004) designed to provide Web access to the vision impaired are becoming increasingly sophisticated in their presentation of complex hypertext information. The more recent arrival of programming languages designed to interface with speech and telephony systems, such as VoiceXML, VoxML, SALT, and Aural style sheets, have improved the level of integration between the Internet (and by virtue the Web) and voice applications (see Goose, Newman, Schmidt & Hue, 2000 for an example). This has seen a rise in over-the-telephone systems delivering hypertext information including email (AudioPoint, 2004), Internet-based forms (AudioPoint, 2004) and Internet voice portals (Tellme, 2004). Researchers have also investigated more novel forms of audible hypertext access, such as Web-TV systems (Braun & Dörner, 1998) and in-vehicle web browsers (Goose & Djennane, 2002).

Despite the rise in audible hypertext systems, few designers have experimented with voice hyperlinks embedded in running text (Balentine & Morgan, 2001). Research to-date has focused on the design of audible hypertext systems making use of a variety of different hyperlink designs using speech and non-speech auditory cues. For example, Morley, Petrie, O'Neill & McNally, (1998) used a high pitch voice preceded by a 'bing' tone to differentiate hyperlinks from surrounding speech. In another study, Asakawa & Itoh, (1998) changed the gender of the voice used to recite hyperlinks. Despite this work, little is known about the effect these different hyperlink designs may have on the listener's ability to comprehend the speech in which they are embedded. No studies have been found that evaluate the relationship between the encoding demands of audible hyperlink perception and their effects on speech comprehension. This gap in the existing work motivated this study.

2. AUDIBLE HYPERLINK DESIGN

Braun et al., (1998) identify seven major components of audible hyperlinks, including the "signal sound", which they describe as the audible representation of the hyperlink that indicates the information to which it links. From an interaction design perspective the function of the hyperlink "signal sound" could be described as providing an acoustic-phonetic signal that

enables the listener to perceive both the hyperlinks presence and the information it refers to with minimal disruption to their understanding of surrounding information. To-date this has been achieved, with varying degrees of success, through the use of speech and non-speech cues.

2.1. Hyperlink "Signal Sounds" Evaluated

Two different hyperlink signal sound designs were evaluated in this study. The first design used a type of speech cue described here as a "voice-change cue", because hyperlink presence is inferred by changes in the characteristics of the speaker's voice used to present hyperlink speech. The second design made use of a non-speech cue followed by a speech cue. The non-speech cue used was an earcon, described by Brewser (2002) as structured combinations of abstract tones representing actions and objects in a user interface. In both designs the style of voice-change cue was male, to differentiate it from the surrounding material, which was spoken in a female voice. The two designs can be described as follows:

1. Voice-change cue (VO): Hyperlink speech recited using a different gender (male) from the surrounding spoken material (female); and
2. Earcon & voice-change cue (EV): Hyperlink speech recited using a different gender (male) from the surrounding spoken material (female) and preceded by an earcon.

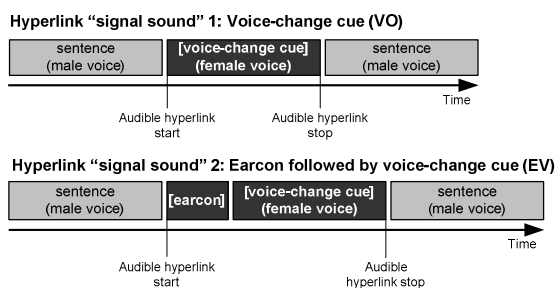Figure 1 illustrates the structures of these two designs.

Figure 1. *Illustration of hyperlink "signal sound" designs evaluated in study*

3. EXPERIMENTAL METHOD

The experiments in this study were conducted using a "sentence verification task", an evaluation method developed by researchers working in comprehension research (see Ralston et al., 1995). During sessions, a test sentence is presented to subjects who must judge whether it is "true" or "false". The truth-value judgments tend to be minor (eg, "birds have wings") and the error rates tend to be very low. The dependent variable of interest tends to be the time it takes for subjects to respond to a given question (response latency) which is used as a measure of comprehension speed (ie, the time it takes the listener to understand and answer the sentence).

The sentence verification task was used in this study because it is able to index response latency to the acoustic-phonetic characteristics of synthetic speech (Ralston at al, 1995), demonstrated by its use in previous experiments to reliably measure the relationship between intelligibility and comprehension of synthetic speech and natural speech (Pisoni

et al., 1997 also see Ralston et al., 1995 for a review of Manaus et al., 1985 and Pisoni et al., 1986). This evidence suggests it may be an appropriate method for evaluating the relationship between the encoding demands of audible hyperlink perception and their effects on speech comprehension.

4. EXPERIMENT 1

The aim of the first experiment was to assess the performance of hyperlink "signal sounds" in terms of their intelligibility and their effect on the comprehension of synthetic sentences. The experimental procedure was adapted from Pisoni et al's., (1987) study of the comprehension of synthetic and natural speech in sentences controlled for intelligibility.

4.1. Hypothesis

The main hypothesis for this experiment was that participants should be able to identify audible hyperlinks embedded in sentences of synthetic speech but that the encoding demands of hyperlink perception would reduce the overall comprehensibility of sentences compared to sentences without hyperlinks.

Hyperlink perception referred to the ease with which subjects would be able to recognise hyperlink speech. The degree of hyperlink intelligibility for each of the designs would provide some insight into the suitability of speech and non-speech cues as "signal sounds" in audible hyperlink design.

Comprehension referred to the speed with which users were able to understand and respond to the truth-value of short sentences of synthetic speech. Given that previous studies demonstrates that the comprehension process of synthetic speech depends on the segmental intelligibility and the difficulty of speech (Ralston et al., 1991, see also, Ralston et al., 1995 for a review) it was possible that differences at both the early stages of perceptual analysis and the later stages of comprehension may impact the same comprehension processes effected by the presence of the auditory cues. To mitigate this possibility, the present study was designed to dissociate effects due to segmental intelligibility and sentence predictability from those related to comprehension processes.

By controlling the level of predictability and intelligibility of the speech, it was hoped that a more direct assessment of the comprehension process associated with the presence of auditory cues would be possible. This would make it possible to draw inferences about processing activities that were not confused with initial differences in sentence intelligibility or difficulty. To accomplish this, sentences of synthetic speech were matched for predictability and intelligibility. Three separate groups of sentences were then developed from the full set of sentences: two groups contained one embedded hyperlink per sentence, each using a different hyperlink design, and a third control group of sentences had no hyperlinks. Each group represented an experimental condition.

The sentence verification task was then used to compare performance of the stimulus materials. If differences in the perception between sentences containing hyperlinks and those without hyperlink are not due only to segmental intelligibility or predictability, then it was expected that there would be differences in response times for a verification task, even if the verification error rates and sentence intelligibilities were comparable for all sentences. This result would demonstrate that the perception and comprehension of sentences containing

embedded audible hyperlinks differs in important ways from the processing of synthetic speech that does not contain hyperlinks.

If sentences that contained audible hyperlinks were more difficult to comprehend than sentences without hyperlinks, then the difference should be influenced by other characteristics of the hyperlinks, such as the type of auditory cue and its configuration within the hyperlink "signal sound". In order to investigate this a sentence verification task was designed to study the difference between a control group of sentences without hyperlinks and two groups of sentences using the different hyperlink designs described. Assuming that people have a limited speech processing capacity, augmenting the voice-change cue with an earcon may increase the resource demands on hyperlink encoding processes. This may, as a consequence, reduce comprehension performance when compared to the hyperlinks that uses only a voice-change cue.

If the hypothesis was correct, then the experiment was expected to yield the following results:

4.1.1. Sentence segmental intelligibility

This refers to the degree of accuracy that subjects were able to recall sentences immediately following presentation and was measured using a transcription task. Low error rates and no significance were anticipated between the intelligibility of sentences across the conditions. This would support the aims of the experiment to present intelligible sentences.

4.1.2. Sentence verification accuracy

This refers to the success of users in comprehending the linguistic content of sentences and was measured by their responses to sentences that were either "true" or "false". Low error rates and no significance were anticipated between the accuracy of responses across the conditions. This would support the aims of the experiment to present intelligible and highly predictable sentences.

4.1.3. Hyperlink intelligibility

This refers to the degree of accuracy with which subjects could recall hyperlinks immediately following presentation and was measured using a transcription task. Low error rates and no significance were anticipated between different hyperlink designs. This would be consistent with the results of previous studies using similar hyperlink designs and support the hypothesis that speech and non-speech auditory cues may be suitable for designing audible hyperlinks.

4.1.4. Sentence verification latency

This refers to the resource demands on comprehension processes and was measured by the lapsed time between sentence presentation and sentence verification. A significant difference was anticipated between response times of sentences including hyperlinks and sentences without hyperlinks. This would be consistent with the hypothesis that hyperlink perception affects sentence comprehension. A significant difference was also anticipated between the response times of the two hyperlink designs. This would be consistent with the hypothesis that different types and configurations of auditory cues effect sentence comprehension.

4.2. Method

4.2.1. Subjects

28 subjects participated in stimuli development and 24 different subjects participated in the experiment. All subjects had UK English as their first language and no history of a speech or hearing disorder. 83% of subjects involved in the experiment had limited or no experience listening to synthetic speech at the time of testing. The remaining 17% were classified as regular listeners.

4.2.2. Materials

4.2.2.1 Stimuli development

There were three stages of stimuli development described below.

4.2.2.1.1 Sentence development

This activity was designed to improve the overall semantic predictability of sentences;

4.2.2.1.2 Sentence intelligibility testing

This activity was designed to ensure that the synthetically produced test sentences had a high level of segmental intelligibility prior to including the audible hyperlinks.

4.2.2.1.3 Hyperlink development

This activity involved identifying words to represent hyperlink speech within each sentence and applying auditory cues to the words in the synthetic speech sound files. Three hyperlink positions and two hyperlink lengths were used to counterbalance their potential effects.

Examples of "high-predictability" test sentences that include a hyperlink are provided in Table 3.1.

TYPE	POSITION	LENGTH	SENTENCES
True	Beginning	1 word	**[Bakers]** make different kinds of bread
True	Beginning	2 word	**[Locked doors]** are unlocked using keys
True	Middle	1 word	Envelopes are used to **[send]** letters
True	Middle	2 word	France is **[a country]** in Europe
True	End	1 word	When it rains people use **[umbrellas]**
True	End	2 word	Vegetarians prefer not to **[eat meat]**
False	Beginning	1 word	**[Torches]** are used when it's light
False	Beginning	2 word	**[People drink]** coffee to stay asleep
False	Middle	1 word	Babies **[cry]** when they are happy

False	Middle	2 word	Jumpers **[are worn]** when it's hot
False	End	1 word	Windows are never made of **[glass]**
False	End	2 word	Prisons are for people **[found innocent]**

Table 3.1. *True and false sentences including hyperlinks*

4.2.2.2 Materials used during experiment

Testing took place in the UCL Interaction Centre usability lab, controlled for sound using a white noise generator. The lab was equipped with a high-quality set of headphones (Somic SM-350) for stimulus playback, a Dell Inspiron 4100 laptop PC used by subjects to provide true/false responses and a set of speakers (Hi-Tex CP-55) to allow the experimenter to monitor stimulus payback during sessions. Stimulus presentation and verification response was controlled and captured using a bespoke system written in Java v1.3.1 and running on the laptop.

4.2.3. Experimental design

24 subjects were tested during the experiment, which followed a one factor within-subjects design with three levels. Sentence group was the within-subjects factor and each sentence group represented a condition. Table 3.2 describes each condition. Stimulus presentation was counterbalanced using a 3x3 Latin square design. Subjects were assigned a condition sequence at random and the sequence of test items presented within each condition was randomised. To ensure that no sentence was repeated across conditions during any of the sessions, the 36 sentences were divided into three groups of 12 test items. An additional three practice trial items were added to each group. Four dependent measures were taken: 1) hyperlink intelligibility 2) sentence segmental intelligibility; 3) sentence verification accuracy; and 4) sentence verification latency.

SENTENCE GROUP	CONDITION
CG	Sentences without hyperlinks
VO	Voice-change cue representing hyperlink speech
EV	Voice-change cue representing hyperlink speech and preceded by an earcon

Table 3.2. *Conditions (1st experiment)*

4.2.4. Procedure

12 sentences were presented to subjects. During each trial, subjects first heard a sentence and then made a forced-choice true/false response. Subjects were instructed to respond as quickly and as accurately as possible when making their true/false decisions. Response latencies were measured using computer-controlled routines from the time a sentence concluded to the moment of the subject's response. After entering their response, subjects were required to transcribe each sentence on a separate printed answer sheet using a pen.

For sentences including hyperlinks, subjects were asked to mark the start and the end points of links within the sentence by placing a "|"before and after the hyperlink speech. This task was included to measure both sentence segmental intelligibility and hyperlink intelligibility.

Subjects completed each exercise in turn following the same procedure. During the course of the experiment the experimenter remained in the room to ensure that subjects responded appropriately. Each session lasted approximately 30 minutes.

At the end of the third exercise subjects completed a post-test questionnaire designed to gather subjective feedback on how easy they felt it was to identify each hyperlink design and overall preference between the two designs. The question formats were a combination of 5-point Likert scales (e.g., ranging from "very easy" to "very" difficult) and free-response. Subjects were given unlimited response time.

4.3. Results

Performance score data was analysed using a one-factor within subjects (repeated measures) ANOVA model. Sentence group was the within-subjects factor (ie, CG, VO, and EV). There were four dependent variables:
1. Sentence segmental intelligibility;
2. Sentence verification accuracy;
3. Link segmental intelligibility; and
4. Sentence verification response latency.

Separate analysis was carried out for each dependent variable to assess the effects of the different sentence groups.

Prior to analysis data was checked for normality. The distribution can be assumed to be normal:
CG: Kolmogorov-Smirnov Z = .410; p = 0.996.
VO: Kolmogorov-Smirnov Z = .954; p = 0.322.
EV: Kolmogorov-Smirnov Z = .542; p = 0.931.

4.3.1. Sentence segmental intelligibility

The error rates for sentence transcription accuracy were very low across all conditions (CG: 0.69%; VO:1.04%; EV:1.39%). This result was consistent with the procedures used to generate sentences of high intelligibility during stimulus development. An analysis of variance on transcription scores to check the effect of the sentence groups on the segmental intelligibility of sentences revealed no significance [$F_{(2,46)}=0.291$, N.S.]. As expected, no differences were found in the immediate recall between synthetic sentences including hyperlinks and those without hyperlinks.

4.3.2. Sentence verification accuracy

Sentence verification error rates were low across all conditions, demonstrating subjects had little difficulty understanding the sentences and carrying out the verification task (CG: 3.47%; VO: 4.17%; EV: 4.51%). Analysis of variance of verification accuracy to check the effect of the sentence groups revealed no significance [$F_{(2,46)}=0.188$, N.S.]. As expected, no differences were found in the true/false accuracy between synthetic sentences including hyperlinks and sentences excluding hyperlinks.

4.3.3. Hyperlink intelligibility

The error rates for hyperlink transcription accuracy were relatively low for sentences across both auditory cue configurations. An analysis of variance on hyperlink transcription scores was performed to check the effect of the different sentence groups revealed no significance [F(1,23)=0.252, N.S.]. Although transcription error rates were slightly higher for VO sentences the difference was not significant.

4.3.4. Sentence verification latency

Response latencies were analysed only for sentences that had been both verified correctly and transcribed correctly. Figure 2 shows the mean verification response latencies for all sentence groups.
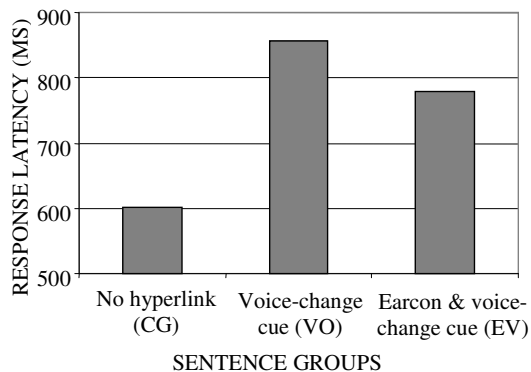

Figure 2. *Mean sentence verification latencies (1st experiment)*

Synthetic sentences not including hyperlinks were consistently responded to more rapidly than synthetic sentences including hyperlinks. The mean difference in response time between the control group (CG) and the sentence groups with hyperlinks was 217ms.

An analysis of variance on sentence verification latency to check the effect of the sentence groups on the time it took subjects to understand the sentences revealed a highly significant effect [F(2,46)=9.478, p<0.001]. Further analysis revealed a highly significant effect between CG and VO [F(1,23)=19.702, p<0.001] and a significant effect between CG and EV [F(1,23)=11.394, p=0.003]. No significance was observed between the two hyperlink sentence groups VO and EV. Results demonstrate that the speed of responding to the control group containing no hyperlinks (CG), was significantly faster than the response times of both sentence groups containing hyperlinks (VO and EV).

4.3.5. Qualitative analysis

4.3.5.1 Ease of audible hyperlink identification

A majority of subjects thought both types of audible hyperlink were either "very easy" or "easy" to identify. A larger proportion of 62.5% rated VO hyperlinks (voice-change cues) "easy" or "very easy" compared to 54.2% for EV hyperlinks (voice-change cues preceded by and earcon). The perception that VO hyperlinks were easier to identify than EV hyperlinks

does not correspond to hyperlink intelligibility scores, where EV error rates were lower than VO by a margin of 0.87% or the or faster mean sentence verification response time for EV sentences by a margin of 79ms.

4.3.5.2 Audible hyperlink preference

A majority of three subjects preferred VO hyperlinks over EV hyperlinks (VO: 12; EV: 9). Three subjects had no preference.

The overall preference for VO hyperlinks is consistent with the perception that they were easier to identify but does not correspond to hyperlink intelligibility and verification latency results.

4.4. Discussion

Results from of the first experiment are discussed with reference to sentence intelligibility and verification; hyperlink intelligibility; and efficiency of sentence comprehension.

4.4.1. Sentence intelligibility & verification

Error rates for both sentence transcription and true/false verification were very low and separate analysis of the dependent variables confirmed no differences in the segmental intelligibility or the true/false accuracy of sentences across all conditions. This suggests:
1. Subjects correctly encoded sentences across all conditions at the time of input; and
2. Subjects successfully comprehended the linguistic content and meaning of the sentences.

The findings are consistent with the stimulus development procedures used to create sentences of "high predictability" and "high intelligibility" before the sentence verification task and confirm the assumption that the types and configurations of auditory cues used to present hyperlinks in this study do not have a significant effect on the encoding or the meaning of short sentences of synthetic speech. These findings are also consistent with previous studies that, despite their different evaluation methods, report no difficulty of subjects understanding speech containing hyperlinks using similar designs (James, 1997; Wynblatt et al., 1997; Morley et al., 1998; Asakawa at al., 1998; Goose et al., 2000).

4.4.2. Hyperlink intelligibility

Error rates for hyperlink transcription were relatively low and analysis confirmed no differences in the segmental intelligibility of hyperlinks across either condition containing hyperlinks (VO: using a voice-change cue and EV: using a voice-change cue and preceded by an earcon), suggesting subjects correctly encoded both hyperlink designs at the time of input.

The findings confirm the assumption that the types and configurations of auditory cues used to present hyperlinks in this study may be suitable in the design of embedded audible hyperlinks, at least in short sentences of synthetic speech. The result is also consistent with previous studies which report no difficulty with subjects identifying hyperlinks of similar design (James, 1997; Wynblatt et al., 1997; Morley et al., 1998; Asakawa at al., 1998; Goose et al., 2000).

4.4.3. Efficiency of sentence comprehension

Response latencies were faster for the control group (CG) containing no hyperlinks than for the two sentence groups containing hyperlinks (VO and EV). A highly significant effect was also observed between the faster response times of the CG group and the slower response times of the VO and EV hyperlink groups. Results demonstrate that sentence verification latency is sensitive to the presence of auditory cues embedded in sentences.

The observed difference cannot easily be attributed to differences in sentence intelligibility due to measures taken to control intelligibility during stimulus development and the low sentence transcription and sentence verification error rates which showed no reliable differences across conditions. Put another way, it appears subjects were able to perceive and encode sentences correctly. They did however have difficulty determining the truth-value of sentences, which required subjects to understand the meaning of sentences and respond appropriately. This is demonstrated in the significant difference in response latency between the control group and the conditions containing hyperlinks. Given these considerations it is suggested that at least part of the observed difference can be attributed to the encoding demands of the auditory cues used to present the hyperlinks.

The suggestion that perception processes compete for the same attentional resources as comprehension processes is consistent with both Kintsch et al's (1978) model of comprehension and Ralston et al's (1995) general assumptions about speech comprehension. It is also consistent with the evidence of Manous et al., (in Ralston et al., 1995) and Luce et al., (1983) that the encoding of synthetic speech incurs greater processing costs than natural speech and that these demands may interact with demands on comprehension resources. This same line of reasoning can be explored further with reference to the "generic verification model of sentence comprehension" put forward by Clark & Chase (in Pisoni et al., 1987) which proposes four stages, each using certain amounts of processing resources:

1. Sentence interpretation;
2. Evaluate relevant external or internal evidence;
3. Compare representations from stages 1 and 2; and
4. Respond with the answer from stage 3.

The sentence is encoded at stage 1 and progressively moves up the processing system to more abstract levels of language processing. If the above model is correct it suggests that acoustic-phonetic interference from the auditory cues during sentence encoding slows the passage of spoken material further up the processing system.

If this assumption is correct it is impossible, based on the current evidence, to determine the precise factors responsible for the increased encoding demands. Are they simply due to acoustic- phonetic interference of the embedded cues or do other factors, such as cue type, cue configuration, cue position and cue length that may also place cognitive demands on limited attentional resources?

The motivation of the second experiment was to investigate these issues further to gain insight into the characteristics of audible hyperlinks that may influence the comprehension process. Prior to the first experiment it was thought that the augmentation of auditory cues may increase demands on hyperlink encoding and reduce sentence comprehension. Contrary to initial expectations, the hyperlink "signal sound" using two auditory cues (a voice-change cue preceded by an earcon) appeared to demand less attentional resources than the hyperlink using one auditory cue (a voice-change cue). Did the presence of an earcon actually improve sentence comprehension? Some subjects also indicated that link position had an effect on their performance. Although post hoc analysis showed no clear indication of the effects of either characteristic, it was felt they deserved further investigation under a more tightly controlled experiment.

5. EXPERIMENT 2

The aim of the second experiment was to examine particular characteristics of audible hyperlinks to measure their specific effect on the comprehension of synthetic sentences. Following a similar approach as the first experiment a sentence verification task was used to study the above effects. The experimental design and procedure used during the first experiment was modified to create a more tightly controlled experiment in which hyperlink characteristics could be examined in closer detail, including reducing the number of conditions under investigation from three to two; adjusting the length of all hyperlinks to one word; and closer monitoring the sentence verification response task to ensure subjects answered questions as quickly as possible.

5.1. Hypothesis

The main hypothesis for this experiment was that that the encoding demands of the different audible hyperlink "signal sounds" and their position within a sentence ("beginning", "middle" and "end") would effect the comprehension performance of sentences in which they appear.

If, as suggested during the first experiment, audible hyperlinks were in some way more difficult to comprehend than sentences without audible hyperlinks, then the difference should be influenced by different characteristics of the hyperlinks, such as the type and configuration of auditory cues used to represent the "signal sound" and the position of the audible hyperlink within the sentence. Assuming that people have a limited speech processing capacity, differences between these characteristics may increase the resource demands on hyperlink encoding processes and, as a consequence, reduce comprehension performance.

5.2. Method

5.2.1. Subjects

A total of 23 subjects took part in the second experiment. 17 were drawn from University College London Department of Psychology's database of subjects and received £5 for their participation. The remaining 6 subjects received no incentive. All subjects had UK English as their first language and no history of a speech or hearing disorder. 80% of subjects had limited or no experience listening to synthetic speech at the time of testing. The remaining 20% were classified as regular listeners.

5.2.2. Materials

5.2.2.1 Stimuli development

Test items from the first experiment were modified to make them suitable for the second experiment. This involved identifying suitable words to represent hyperlink speech and applying the appropriate auditory cues to represent the hyperlink "signal sound". Hyperlink speech was selected according to the same criteria used in the first experiment, however, on this occasion only one word links were generated. Two groups of 36 test items were produced: each group representing one of the hyperlink designs.

5.2.2.2 Materials used for experiment

Materials were the same as those used in the first experiment.

5.2.3. Experimental design

23 subjects were tested during the experiment, which followed a two factor within-subjects design with three levels per factor. Hyperlink design and hyperlink position were the within-subjects factors. Table 3 describes the conditions used in the experiment.

HYPERLIN DESIGN	HYPERLINK POSITION
VO: *Voice-change cue representing hyperlink speech*	Beginning
	Middle
	End
EV: *Voice-change cue representing hyperlink speech and preceded by an earcon*	Beginning
	Middle
	End

Table 3 *Conditions (2nd experiment)*

Stimulus presentation was counterbalanced using a 2x2 Latin square design. As with the first experiment, subjects were assigned a condition sequence at random and the sequence of test items presented within each condition was randomised. To ensure that no sentence was repeated across conditions during any of the sessions, the 36 sentences were divided into two groups of 18 test items. An additional three practice trial items were added to each group. As with the first experiment, four dependent measures were taken: 1) hyperlink intelligibility 2) sentence segmental intelligibility; 3) sentence verification accuracy; and 4) sentence verification latency.

5.2.4. Procedure

The same procedures used during the first experiment were followed with one exception. During instruction and training the experimenter placed a greater emphasis on the goal of subjects to answer questions as quickly as possible. This was designed to provide more tightly controlled measurements between the two sentence groups. Each session lasted approximately 30 minutes.

5.3. Results

The results reported below were based on data collected from 20 subjects. Data from three male subjects were removed prior to analysis.

Performance score data was analysed using a two-factor within subjects (repeated measures) ANOVA model. Hyperlink design and hyperlink position were the within-subjects factors. Consistent with the first experiment, there were four dependent variables: 1) sentence segmental intelligibility, 2) sentence verification accuracy, 3) hyperlink segmental intelligibility, and 4) sentence verification response latency. Separate analysis was carried out for each dependent variable to assess the effects of the different hyperlink designs. The effect of hyperlink position was only analysed for sentence verification response latency.

Prior to analysis data was checked for normality. The distribution can be assumed to be normal:
VO: Kolmogorov-Smirnov Z = .663; p = 0.771.
EV: Kolmogorov-Smirnov Z = .576; p = 0.894.

5.3.1. Sentence segmental intelligibility

As with the first experiment, the error rates for sentence transcription accuracy were very low (0.83% for both conditions). An analysis of variance on transcription scores to check the effect of hyperlink design on the segmental intelligibility of sentences revealed no significant effect [F(1,18)=0.000, N.S.]. These results confirm the findings of the first experiment suggesting subjects correctly encoded sentences across all conditions at the time of input

5.3.2. Sentence verification accuracy

Also consistent with the first experiment, sentence verification error rates were low across both conditions (VO: 5.00%; EV: 4.17%) confirming subjects had little difficulty in understanding the sentences and carrying out the verification task. Consistent with results from the first experiment, analysis of the variance of verification accuracy revealed no significant effect of hyperlink design [F(1,18)=0.81, N.S.], suggesting subjects successfully comprehended the linguistic content and meaning of the sentences.

5.3.3. Hyperlink segmental intelligibility

Also consistent with the first experiment, the error rates for hyperlink transcription accuracy were low for both hyperlinks.

An analysis of variance on hyperlink transcription scores revealed no significant effect of hyperlink design [F(1,18)=0.213, N.S.]. Consistent with the first experiment, VO hyperlinks had a higher rate of error than EV hyperlinks. These results add evidence to the findings from the first experiment that subjects correctly encoded the words represented by the hyperlinks at the time of input.

5.3.4. Sentence verification response latency

As with the first experiment, response latencies were analysed only for sentences that had been both verified correctly and transcribed correctly. Figure 3 shows the mean verification response latencies for sentences across both conditions.
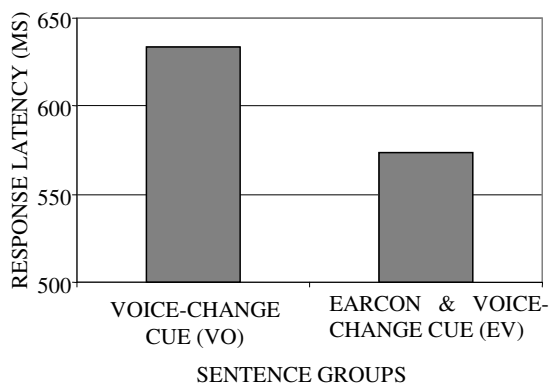
Figure 4.2. *Mean sentence verification latencies (2nd experiment)*

A consistent effect of auditory cue can be observed in the response times for both experiments. Synthetic sentences with the EV hyperlink were responded to more rapidly than those with the VO hyperlink. Mean response times for the second experiment were quicker for both conditions than for the first experiment, possibly due to stricter monitoring of subject response times mentioned.

An analysis of variance on sentence verification response latency to check the effect of hyperlink design on the comprehensibility of sentences revealed a significant effect [$F(1,19)=4.681$, $p=0.043$]. A significant difference was found in the response latencies between sentences with the VO hyperlink and those with the EV hyperlink.

An analysis of the effect of auditory cue position (i.e. "beginning", "middle" and "end") on sentence verification response latency failed to reach significance.

5.3.5. Qualitative analysis

5.3.5.1 Ease of audible hyperlink identification

Consistent with feedback from the first experiment a majority of subjects thought both types of audible hyperlink were either "very easy" or "easy" to identify. Also consistent with the previous findings, a larger proportion of 90% rated VO hyperlinks (voice-change cues) "easy" or "very easy" compared to 85% for EV hyperlinks (voice-change cues preceded by and earcon). Also consistent with the previous evidence, these perceptions did not correspond to the higher performance of the EV hyperlinks in terms of hyperlink intelligibility error rates (by a margin of 0.7%) or faster mean sentence verification response time for EV sentences by a margin of 60ms.

5.3.5.2 Audible hyperlink preference

Consistent with the first experiment, a majority of three subjects preferred VO hyperlinks over EV hyperlinks (nine subjects preferred VO and seven preferred EV). Four subjects had no preference. The preference for VO hyperlinks is consistent with both the first experiment and the perception that they were easier to identify but does not correspond to the hyperlinks intelligibility and verification latency results.

5.4. Discussion

Results from of the second experiment are discussed with reference to three issues:
1. Sentence intelligibility and verification;
2. Hyperlink intelligibility; and
3. Efficiency of sentence comprehension.

5.4.1. Sentence intelligibility & comprehension

Consistent with the results of the first experiment, error rates for both sentence transcription and true/false verification were very low and separate analysis of the hyperlink designs confirmed no differences in the segmental intelligibility or the true/false accuracy of sentences across either condition. The evidence supports the conclusions of the first experiment that subjects correctly encoded and successfully comprehended the meaning of sentences across both conditions. This confirmed the assumption that the types and configurations of auditory cues used to present hyperlinks did not have a significant effect on the encoding or the meaning of the sentences.

5.4.2. Hyperlink intelligibility

Also consistent with the results of the first experiment, error rates for hyperlink transcription were very low and analysis confirmed no differences in the segmental intelligibility of hyperlinks across either hyperlink design condition, suggesting subjects correctly encoded both hyperlink designs at the time of input. The evidence confirms the assumption that the types and configurations of auditory cues used to present hyperlinks in this study may be suitable in the design of embedded audible hyperlink in isolated sentences of synthetic speech.

5.4.3. Efficiency of sentence comprehension

Response latencies between hyperlink conditions was quicker for EV (earcon followed by voice-change) than for VO (voice-change only). This is consistent with the trend observed in the results of the first experiment. Furthermore, a highly significant effect was observed between response latencies, demonstrating that verification latency is sensitive to different types and configurations of auditory cues. The mean difference in response latency was 60ms. This margin should be noted with reference to the fact that sentences with EV hyperlinks were 500ms longer than sentences with VO hyperlinks. It appears that at least part of this difference can be attributed to different encoding demands of the auditory cues used to present the audible hyperlinks. In addition, no effect of response latency was observed for hyperlink position. Each of these findings will be considered in turn.

The results demonstrate that verification latency is sensitive to the different hyperlink designs, suggesting the speed and efficiency of sentence comprehension varies for different types of cue and cue configurations. Response times were quicker for hyperlinks using an earcon preceded by a voice-change cue than those using only a voice-change cue. Despite subjective feedback suggesting the earcon distracted subjects during the task, results demonstrate that it actually improved task performance. It appears that preceding a hyperlink with a short, non-speech cue, such as an earcon, reduces the encoding demands of hyperlink perception and in doing so improves sentence comprehension. This is supported by the suggestion of Ralston et al., (1995) that in certain circumstances subjects may

reallocate spare resources from acoustic-phonetic encoding of synthetic speech to more abstract cognitive processes.

Turning to the properties of the earcon itself, one could speculate that its presence at the start of the hyperlink may have alerted the user to its presence before the onset of the hyperlink text. This is consistent with the post-session feedback of some subjects who identified the "prompt" and "attention grabbing" qualities of the earcon. It is also supported by Brewster et al., (2002), who point out both the attention grabbing qualities and expressive capabilities of non-speech sounds in graphic and auditory user interfaces. It is possible that such an alerting effect reduces the encoding demands of hyperlink perception which, in the case of the voice-only hyperlinks, is less abrupt and begins at the point of word recital. This line of thinking is supported by Brewster's (1993) findings which suggest non-speech sounds are an effective means of communicating information in auditory user interfaces.

As mentioned, contrary to expectation no effect on sentence verification latency was observed for link position. This may simply indicate that hyperlink position does indeed have no discernable effect on the encoding demands of hyperlink perception in synthetic sentences. However, it is possible that this result was due to other factors that were not detectable within the experimental design. One explanation may be that one-word links embedded in six-word sentences are too insensitive to detect any difference in cognitive demands required to encode the links located at different positions within the speech. If this line of reasoning is correct, then a similar study that makes use of longer sentences or passages of fluent speech or that evaluates hyperlinks using more than one word may yield results. This could be one area for future study.

6.   GENERAL DISCUSSION

Four main findings can be drawn from the result of these experiments:
1. Hyperlink designs did not significantly reduce sentence intelligibility or sentence meaning.
2. Hyperlink designs did not significantly reduce hyperlink intelligibility.
3. Hyperlink "signal sounds" significantly reduced the comprehension of sentences of synthetic speech compared to sentences without hyperlinks.
4. There was a significant difference in comprehension performance between different "signal sound" designs.
5. There was a tendency for subjects to prefer VO hyperlinks and perceive them as easier to identify than EV hyperlinks.

These findings may have implications for the design of audible hyperlinks and the suitability of audible hypertext systems in certain high-workload task environments.

7.   CONCLUSION

In summary, the results of this study demonstrate that speech and non-speech cues can be perceived as audible hyperlinks in isolated sentences of synthetic speech. This supports the findings of previous studies and suggests that auditory cues may be effective in the design of audible hyperlinks. However, some aspects of the comprehension process, possibly the encoding of the initial representation, are affected by the acoustic-phonetic characteristics of the audible hyperlinks within the speech signal. Results also demonstrate that different acoustic-phonetic characteristics of audible hyperlinks, represented using different types and configurations of speech and non-speech cues, affect the comprehension process. Using short, meaningful sentences that were controlled for segmental intelligibility and predictability, verification latencies were found to be reliably shorter for sentences of synthetic speech without hyperlinks than sentences of synthetic speech that included hyperlinks. Sentence verification latencies were also found to be reliably shorter for hyperlinks that preceded a voice-change cue with an earcon than those that made use of only a voice-change cue. These results may have implications for the design of audible hyperlinks which may extend to an evaluation of the suitability of audible hypertext to support a given task or context of use, particularly those in high-workload and high-information situations.

8.   REFERENCES

[1] C. Asakawa and T. Itoh, (1998). *User interface of a home page reader*. Proceedings ACM Conference on Assistive Technologies (ASSETS) 1998.

[2] AudioPoint, (2004). *Email by Phone*. AudioPoint website: http://www.myaudiopoint.com/EMbP_1pagePDF.pdf, (visited: 08/2004).

[3] AudioPoint, (2004). *VoiceForms*. AudioPoint website: http://www.myaudiopoint.com/voice_form.pdf, (visited: 08/2004).

[4] B. Balentine and D. Morgan, (2001). *How to Build a Speech Recognition Application: A Style Guide for Telephony Dialogues*. EIG Press, San Ramon, California.

[5] N. Braun and R. Dörner, (1998). *Using Sonic Hyperlinks in Web-TV*. Proceedings of the Fifth International Conference on Auditory Displays (ICAD'98), Glasgow.

[6] S.A. Brewster, (2002). *Chapter 12: Non-speech auditory output*. J. Jacko and A. Sears (Eds.), The Human Computer Interaction Handbook pp. 220-239. Lawrence Erlbaum Associates, USA.

[7] Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). *An evaluation of earcons for use in auditory human-computer interfaces*. In Ashlund, Mullet, Henderson, Hollnagel & White (Eds.), Proceedings of ACM/IFIP INTERCHI'93, (pp. 222-227), Amsterdam: ACM Press, Addison-Wesley.

[8] S. Goose, M. Newman, C. Schmidt and L. Hue, (2000). *Enhancing Web Accessibility Via the Vox Portal and a Web Hosted Dynamic HTML<->VoxML Converter*. Proceedings International World Wide Web, Amsterdam, 2000.

[9] S. Goose and S. Djennane, (2002). *WIRE^3: Driving around the Information Super-Highway*. Personal and Ubiquitous Computing **6**, 164-175.

[10] IBM, (2004). *IBM Accessibility Center: IBM Home Page Reader 3.0*. IBM website: http://www-3.ibm.com/able/solution_offerings/hpr.html, (visited 08/2004).

[11] F. James, (1996). *Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext* in Proceedings of the International Conference on Auditory Display (ICAD), pp 97-103.

[12] JAWS (2004). *JAWS for Windows*. Freedom Scientific website: http://www.freedomscientific.com/fs_products/software_jaws.asp, (visited: 08/2004).

[13] A. King, G. Evans, and P. Blenkhorn, (2004). *WebbIE: a web browser for visually impaired people*. S. Keates, P. J.

Clarkson, P. Langdon, and P. Robinson (Eds), Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology, pp. 35-44. Springer-Verlag, London, UK.

[14] W. Kintsch and T. A. van Dijk, (1978). *Toward a model of text comprehension and production* in Psychological Review **85**, 363-394.

[15] P. A. Luce, T. C. Feustel and D. B. Pisoni, (1983). *Capacity demands in short-term memory for synthetic and natural speech.* Human Factors **35**, 17-32.

[16] S. Morley, H. Petrie, A. O'Neill and P. McNally, (1998). *Auditory Navigation in Hyperspace: Design and Evaluation of a Non-Visual Hypermedia System for Blind Users*. Proceedings of ASSETS '98, the Third Annual ACM Conference on Assistive Technologies, Los Angeles, CA.

[17] D.B. Pisoni, L.M. Manous and M.J. Dedina, (1987). *Comprehension of natural and synthetic speech: effects of predictability on the verification of sentences controlled for intelligibility.* Computer Speech and Language **2**, 303-320.

[18] J.V. Ralston, D.B. Pisoni, S.E. Lively, B.G. Greene and J.W Mullennix, (1991). *Comprehension of Synthetic Speech Produced by Rule: Word Monitoring and Sentence-by-Sentence Listening Times.* Human Factors 33(4), 471-491.

[19] J.V. Ralston, D.B. Pisoni and J.W. Muliennix, (1995). *Perception and comprehension of speech.* A.K. Syrdal, R.W. Bennett, S.L.Greenspan (Eds.), Applied Speech Technology, pp. 233-288. Boca Raton: CRC Press.

[20] Tellme, (2004). *About us: At a glance.* Tellme Networks website: http://www.tellme.com/about.html (visited: 08/2004).

[21] M. Wynblatt, D. Benson and A. Hsu, (1997). *Browsing the World Wide Web in a Non-Visual Environment*. Proceedings of the International Conference on Auditory Display (ICAD), pp. 135-138, November, 1997.