CUEING HYPERLINKS IN AUDITORY INTERFACES

Research paper for the ICAD05 workshop "Combining Speech and Sound in the User Interface"

Dan Hamer-Hodges, Simon Y. W. Li and Paul Cairns

University College London Interaction Centre,
University College London
London, WC1E 7DP, 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

Audible hypertext content is becoming increasingly available in commercial desktop applications and over-the-telephone systems. Voice browsers [1] and screen-readers [2] designed to provide Web access to the visually 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 [3]. This has seen a rise in over-the-telephone systems delivering hypertext information including email [4], Internet-based forms [5] and Internet voice portals [6]. Researchers have also investigated more novel forms of audible hypertext access, such as Web-TV systems [7] and in-vehicle web browsers [8].

Despite the rise in audible hypertext systems, few designers have experimented with voice hyperlinks embedded in running text [9]. Barring a few exceptions, [7] and [10], studies to-date have tended to focus on the development of interactive systems rather than on the design and evaluation of audible hyperlinks. Studies have made use of a variety of audible hyperlink designs using speech and non-speech auditory cues. For example, Morley, Petrie, O'Neill and McNally [11] used a high pitch voice preceded by a 'bing' tone to differentiate hyperlinks from surrounding speech. In another study, Asakawa & Itoh [12] changed the gender of the voice used to recite hyperlinks. Despite this work, there is little experimental evidence available to judge the performance characteristics of different hyperlink designs in terms of intelligibility and comprehension. One exception is Susini, Vieillard, Deruty, Smith and Marin's [10] evaluation of audible hyperlinks, which suggests that although all of the sounds evaluated were successfully identified by subjects, narrow band sounds were significantly more effective and sounds in the 1-3kHz spectral range had a significant effect on "nuisance value". 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 work.

Two studies were performed to evaluate the effect of different types of cues for audible hyperlinks. In a controlled study of this sort, it is not possible to measure the full range of possible hyperlink designs so our choice of cues is discussed in the next section followed by an explanation of the experimental method used to evaluate them. The experiments show that audible hyperlink cues do have a clear effect on sentence comprehension, even when sentences are intelligible and predictable. Even so, hyperlinks are easily recognised and users seem to find them acceptable.

2. HYPERLINK DESIGNS EVALUATED

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 [13] [14][15][16].

This provides interesting constraints on what would constitute useful and applicable cues for hyperlinks. Clearly, the cues had to be suitable for auditory displays – designs for visual displays, such as the auditory cues used to provide feedback about the probable type of information at the other end of a hyperlink before activation, used by Albers and Bergman [17] would not be suitable. Earcons are natural choices of non-speech cue but they could not be used in isolation as they have been shown not to be effective [18] and also confusable with other auditory warnings [18]. Accordingly, the primary cue was a change in speaker voice reciting the hyperlink speech. This distinct change in voice was intended to be similar to IBM's Home Page Reader [1] while avoiding the risk of voice distortion that can occur by simply changing the pitch of the speaker's voice [11].

The two final hyperlink designs evaluated in this study were:

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.

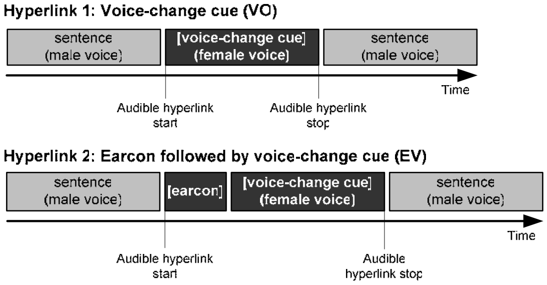

Figure 1. *Illustration of Hyperlink "Signal Sound" Designs Evaluated in Study*

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 [16]. 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 demonstrated by its use in previous experiments to reliably measure the relationship between intelligibility and comprehension of synthetic speech and natural speech [16]. On this basis it was decided to 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 the hyperlink designs in terms of their intelligibility and their effect on the comprehension of synthetic sentences. The experimental procedure was adapted from Pisoni, Manous and Dedina's study of the comprehension of synthetic and natural speech in sentences controlled for intelligibility [14].

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 demonstrate that the comprehension process of synthetic speech depends on the segmental intelligibility and the difficulty of speech [15][16] 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: two groups contained either one embedded VO or EV hyperlink per sentence and a third control group contained no hyperlinks. Each group represented an experimental condition.

If differences in the perception and comprehension between sentences containing hyperlinks and those without hyperlinks 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. These differences should be influenced by characteristics of the hyperlinks, such as the type of auditory cue and its configuration within the hyperlink design. 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 use 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. Low error rates and no significance were anticipated between the intelligibility of sentences across the conditions.

4.1.2. Sentence Verification Accuracy

This refers to the success of users in comprehending the truth-value of sentences. Low error rates and no significance were anticipated between the accuracy of responses across the conditions.

4.1.3. Hyperlink Intelligibility

This refers to the degree of accuracy with which subjects could recall hyperlinks immediately following presentation. Low error rates and no significance were anticipated between different hyperlink designs.

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. A significant difference was also anticipated between the response times of the two hyperlink designs.

4.2. Method

4.2.1. Subjects (stimuli development)

28 subjects participated in stimuli development phase of the experiment. All subjects had UK English as their first language and no history of a speech or hearing disorder.

4.2.1.1 Stimuli Development

The stimuli were short sentences of synthetic speech recorded as audio files in mp3 format (128kbps 48.00 kHz) using Text Aloud MP3 version 1.4 text-to-speech system with AT&T Natural Voices (Charles and Audrey - UK English). All sentences were six words long deliberately devised to be intelligible and predictable before the addition of hyperlinks. This was to ensure that any effect of the audible cues was unlikely to be confounded with sentence length, sentence comprehension or sentence truth-value. Hyperlinks were added to the sentences in three positions: beginning, middle or end. The length of the hyperlink was either one or two words. The EV hyperlink design used the same voice-change cue but was preceded by an earcon. The earcon sound was based on the 'Delete' earcon available in Brewster's Hyper Card stack [19], which was edited down from 700ms to 500ms. Examples test sentences are shown in Table 1.

TYPE	POSITION	LENGTH	SENTENCES
True	Beginning	1 word	**[Bakers]** make different kinds of bread
True	Middle	2 word	France is **[a country]** in Europe
True	End	1 word	When it rains people use **[umbrellas]**
False	Beginning	2 word	**[People drink]** coffee to stay asleep
False	Middle	1 word	Babies **[cry]** when they are happy
False	End	2 word	Prisons are for people **[found innocent]**

Table 1. *True and False Sentences Including Hyperlinks*

4.2.1.2 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.

4.2.2. Experimental Design

The design of the experiment consisted of a single factor (sentence group) with three levels; CG, VO and EV. Sentence group was the within-subjects factor and each sentence group represented a condition. Table 2 describes each condition.

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 2. *Conditions (1st experiment)*

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. Sentence segmental intelligibility;
2. Sentence verification accuracy;
   Hyperlink intelligibility; and
3. Sentence verification latency.

4.2.3. Subjects (main experiment)

24 subjects participated in the testing phase of the experiment. 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.

4.2.4. 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 repeated ANOVA. Data was confirmed to be normal using the Kolmogorov-Smirnov test. 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. 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 sentence groups.

4.3.1. Sentence Intelligibility, Sentence Verification and Hyperlink Intelligibility

The error rates for sentence transcription accuracy, sentence verification accuracy and hyperlink intelligibility were very low across all conditions. Hyperlink intelligibility was slightly higher for EV sentences than VO sentences. An ANOVA to check the effect of the sentence groups on sentence transcription accuracy, sentence verification accuracy and hyperlink intelligibility revealed no significant differences.

4.3.2. Sentence verification response 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.

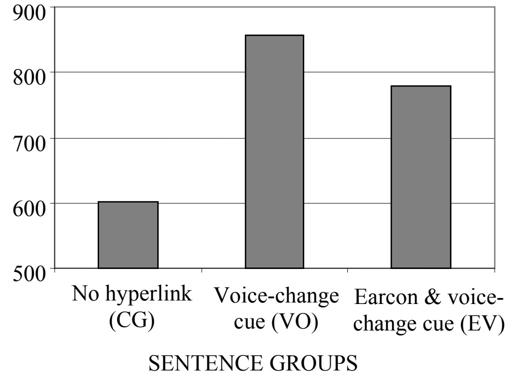

Figure 2. *Mean Sentence Verification Latencies (1st Experiment)*

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]. Planned comparisons 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).

4.3.3. Qualitative Analysis

4.3.3.1 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 “easy” or “very easy” compared to 54.2% for EV hyperlinks. 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 faster mean sentence verification response time for EV sentences by a margin of 79ms.

4.3.3.2 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.

Findings will then be related to theoretical models of comprehension.

4.4.1. Sentence Intelligibility, Sentence Verification and Hyperlink Intelligibility

Error rates for sentence transcription, true/false verification and hyperlink transcription were very low and separate analysis of the dependent variables confirmed no differences across all conditions. This suggests:

1. Subjects correctly encoded sentences across all conditions at the time of input;
2. Subjects successfully comprehended the linguistic content and meaning of the sentences; and
3. 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.

This result was consistent with the procedures used to generate sentences of high intelligibility and predictability during stimulus development. It is also consistent with previous studies which report subjects had no difficulty identifying hyperlinks that used similar designs [3][11][12][18][20].

4.4.2. 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 that 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. This suggests that at least part of the observed difference can be attributed to the encoding demands of the auditory cues used to present the hyperlinks.

4.4.3. Encoding Demands & Models of Comprehension

The suggestion that perception processes compete for the same attentional resources as comprehension processes is consistent with both Kintsch and van Dijk's model of comprehension [21] and Ralston et al.'s general assumptions about speech comprehension [16]. It is also consistent with evidence that the encoding of synthetic speech incurs greater processing costs than natural speech and that these demands may interact with demands on comprehension resources [13]. This same line of reasoning can be explored further with reference to the "generic verification model of sentence comprehension" put forward by Clark & Chase [14] 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 moves up the 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 particular 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 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 EV hyperlink using two auditory cues (a voice-change cue preceded by an earcon) appeared to demand less attentional resources than the VO hyperlink which used only 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. This included reducing the number of conditions under investigation from three to two; adjusting the length of all hyperlinks to one word; and closer monitoring of 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 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 GBP £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

The experiment was a 3 × 3 within-subject design. Hyperlink design and hyperlink position were the within-subjects factors. Table 3 describes the conditions used in the experiment.

HYPERLINK 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. The same dependent measures used during the first experiment were also taken:

1. Sentence segmental intelligibility;
2. Sentence verification accuracy;
3. Hyperlink segmental intelligibility; and
4. Sentence verification response 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

Data from three male subjects were removed prior to analysis resulting in data from 20 subjects in total. One subject had a hearing impairment which became apparent to the experimenter during the session. The other two subjects did not have UK English as a first language.

Performance score data was analysed using a two-factor repeated ANOVA. Again, normality was confirmed using the Kolmogorov-Smirnov test. Hyperlink design and hyperlink position were the within-subjects factors. 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.

5.3.1. Sentence Intelligibility, Sentence Verification and Hyperlink Intelligibility

As with the first experiment the error rates for sentence transcription accuracy, sentence verification accuracy and hyperlink intelligibility were very low across all conditions. Also consistent with the first experiment, VO hyperlinks had a higher rate of error than EV hyperlinks. An ANOVA to check the effect of the sentence groups on each dependent variable revealed no significance. These results confirm the findings of the first experiment.

5.3.2. 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.

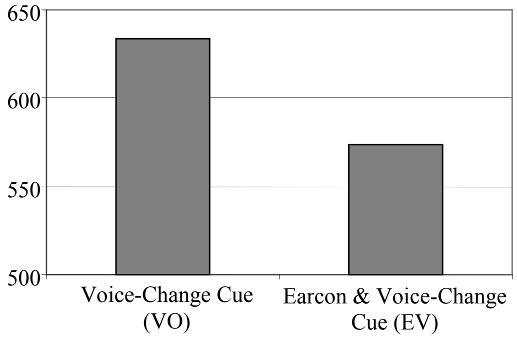

Figure 3. *Mean Sentence Verification Latencies (2nd Experiment)*

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]. Thus 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.

5.3.3. Qualitative analysis

5.3.3.1 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 lower rate of hyperlink transcription errors for EV hyperlinks by a margin of 0.7% or faster mean sentence verification response time for EV sentences by a margin of 60ms.

5.3.3.2 Audible hyperlink preference

Consistent with the first experiment, a majority of 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 compared to EV hyperlinks.

5.4. Discussion

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

5.4.1. Sentence Intelligibility, Sentence Verification and Hyperlink Intelligibility

Consistent with the results of the first experiment, error rates sentence transcription, true/false verification and hyperlink intelligibility were very low and separate analysis of variance revealed no significance. The evidence supports the conclusions of the first experiment that subjects correctly encoded and successfully comprehended the meaning of sentences across both conditions. The evidence also 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 hyperlinks in isolated sentences of synthetic speech.

5.4.2. 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 Ralston et al.'s [16] suggestion of 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 speech. This is consistent with the post-session feedback of some subjects who identified the "prompt" and "attention grabbing" qualities of the earcon. 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 hyperlink speech recital. This explanation is supported by Brewster's findings which suggest non-speech sounds are an effective means of communicating information in auditory user interfaces [22].

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. CONCLUSION

Our studies show that audible hyperlinks do have an effect on the ability of people to comprehend sentences as reflected in verification latencies. This is even in the situation where the sentences are highly predictable and intelligible. Despite this, and in agreement with previous studies, speech cues with or without non-speech cues can be effective in the design of audible hyperlinks.

Additionally, the use of a non-speech cue together with a speech cue does give a measurable improvement in sentence verification latencies. This seems to be at the expense of subjective user preference though this is not supported statistically and it would be worth investigating further to see if users reliably prefer not to have non-speech cues.

These results may have implications for the design of audible hyperlinks though clearly, there are many more parameters such as types of speech and non-speech cue and integration with existing auditory displays. In addition, there is the question of to what extent the experimental findings relate to the use of audible hypertext to support a given task in a specific context of use, particularly those in high-workload and high-information situations.

7. REFERENCES

[1] 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).

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

[3] 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.

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

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

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

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

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

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

[10] P Susini, S Vieillard, E Deruty, B Smith & C Marin, (2002). Sound navigation: Sonified hyperlinks. Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July, 2002.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] M.C. Albers & E. Bergman, (1995). *The Audible Web: Auditory Enhancements for Mosaic*. Proceedings of CHI 95, pp. 318-319.

[18] 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.

[19] Brewster, S.A., (1998). *Earcon Experiments.* Multimodal Interaction Group website: http://www.dcs.gla.ac.uk/~stephen/earconexperiment1/earcon_expts_1.shtml (visited 07/2004).

[20] 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.

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

[22] 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.