Improving Communication of Visual Signals by Text-to-Speech Software

Robert F. Lorch, Jr.[1] and Julie Lemarié[2]

[1] University of Kentucky
Lexington, Kentucky, USA
rlorch@email.uky.edu
[2] Université de Toulouse – Le Mirail, CLLE-CNRS
Toulouse, France
lemarie@univ-tlse2.fr

Abstract. Printed signals are well-documented aids to reading and memory for expository text. Despite their usefulness, many TTS applications fail to adequately communicate signaling. A theoretical framework called "SARA" provides an analysis of printed signals that identifies what specific information should be rendered in order to preserve the signaling function. Further, SARA identifies two important criteria – the availability criterion and the accessibility criterion – that should help guide the evaluation of alternative auditory renderings of signals.

1 Introduction

If a text consists of continuous prose (e.g., narrative), text-to-speech (TTS) software is very good at producing a clearly intelligible auditory rendering of the text [1-2]. However, much text contains non-sentential content: A set of instructions is likely to contain illustrations of the procedures to be carried out. A scientific article is likely to contain tables and graphs. Exposition of all types frequently contains various signaling devices (e.g., headings, typographical contrast) used by the author to emphasize specific content or to highlight aspects of the organization of a text. These non-prose sources of information in a text pose major challenges to attempts to render printed text in other modalities [1], [3-4]. It seems readily apparent that visual, analog nature of graphs presents obstacles to the goal of effectively communicating their content via a nonvisual representation. It may be less apparent that signaling devices such as headings and typographical contrast pose any problems for TTS, but they do [3], [5].

This article focuses on the question of how TTS applications might effectively communicate signaling devices in printed text. We organize our presentation as follows: First, we define signaling devices and briefly review the psychological literature documenting their influence on processing of expository text. Second, we describe our theoretical framework for analyzing signaling devices. We present our framework as an answer to the question of what information must TTS software communicate to effectively render a printed signaling device. Finally, we consider the other important question that must be answered if printed signals are to be effectively

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part III, LNCS 8011, pp. 364–371, 2013.

communicated by TTS software; namely, what forms of auditory representation might be used to effectively communicate visually signaled information?

2 Effects of Signals on Text-Processing

Understanding technical information or following the exposition of a complex argument are examples of common challenges faced by citizens in an educated society. Part of the challenge for a reader or listener is following the organization of an exposition, and part of the challenge is distinguishing more important from less important content. Authors often attempt to help their audiences by using signaling devices to highlight text organization and to emphasize important content [6]. In printed text, signals include writing devices such as: headings, outlines, topical overviews, typographical contrast, preview sentences, summaries and importance indicators.

There is an extensive literature in cognitive and educational psychology that documents a range of ways in which signaling devices support text-processing [7]. Signals that emphasize specific text content produce more attention and better memory for the signaled content. Specifically, the use of typographical contrast (e.g., boldface, italics, color) is effective at attracting attention to the signaled content [8] and improving memory for the content [8-10]. Similarly, the use of numbers to highlight a list results in more attention and better memory for the content [11].

Various signaling devices may be used singly or in combination to highlight text organization to the benefit of readers. Numbering a list of items produces better memory for the organization of the content [11]. Providing an overview of text topics speeds reading of sentences that introduce the topics later in the text [12-13] and improves recall of content [14] by aiding memory for the text topics [15-16]. Preview and review sentences that make connections between topics that are distant in the surface structure of a text result in greater attention to those connections and better memory for them after reading [17]. Headings, and headings in combination with topical overviews and summaries, have been repeatedly shown to support online processing and subsequent recall of text. The presence of headings in a text affects readers' calibrations of their progressive eye movements [18-19] and how they scan a text when searching it for the answers to questions [20]. Headings can facilitate the search for information in a text [21-22] and they serve as reference points for readers when they are reading to summarize a text [23-24]. In fact, the presence of organizational signals in a text can alter readers' text-processing strategies [25]. In the absence of signaling of text organization, even mature readers often construct a very linear representation of text structure that is minimally organized around the major text topics [16], [26-27]. When signaling devices emphasize text organization, readers construct a hierarchical representation of text structure that focuses on the major text topics [16], [26-27]. This change in text-processing strategy and the resulting effects on the nature of the text representation leads to better memory for the text topics and their organization and that, in turn, usually leads to better overall memory for text content [16], [26], [28].

Signaling devices are generally found to facilitate text processing but there is evidence that the extent of their benefits depend on both text and reader characteristics. Organizational signals are increasingly beneficial as text organization becomes more

complex [28] or more difficult to perceive [15]. Readers who lack background knowledge relevant to understanding a text appear to depend more on signals to text organization [29] and content [30-31]. Finally, there is some evidence to suggest that poorer readers benefit more from the presence of signals than better readers [32].

To sum up, signaling devices are ubiquitous in expository texts. Authors use them to guide readers in their efforts to understand text organization and to sort important from less important content. The empirical literature demonstrates that signaling devices do, in fact, serve these purposes well. Given this state of affairs, it is desirable that printed signals be effectively communicated by TTS applications. To do this, two problems must be solved: (1) the information carried by printed signals must be identified and (2) effective means of auditorily communicating that information must be identified.

3 SARA: A Theoretical Framework for Analyzing Signaling Devices

The problem of identifying the information carried by printed signals is addressed by a theoretical framework we call "SARA", which stands for "signaling available, relevant, accessible" information [33]. The acronym stands for the three criteria that we hypothesize must be met in order for a signaling device to influence text processing: The signal must code information (i.e., make "available") in a way that is easy for the reader to use (i.e., "accessible") and the information must be pertinent (i.e., "relevant") to the reader's goals. The criterion of "availability" is our concern in the present context; that is, what distinct types of information does a given signal encode? We propose that 7 distinct types of information may be identified:

1. **Demarcation:** Some signaling devices indicate major boundaries in the structure of a text. This information is often communicated in a purely visual way (e.g., a line return).
2. **Emphasis:** Some signaling devices provide emphasis to text content, often by purely visual means (e.g., typographical contrast).
3. **Hierarchical Organization:** Outlines, topical overviews, systems of headings all typically carry information about the hierarchical organization of topics or themes within a text. This information can be coded verbally but it is also often communicated visually by variation in typography and spacing.
4. **Linear Organization:** Sequencing of related text sections is often signaled by numbering but it can also be communicated nonverbally (e.g., by bullets)
5. **Labeling**: Signals such as headings can be used to provide verbal labels for text sections. The labels facilitate reference to specific sections by the author or reader.
6. **Topic Identification:** Signals such as headings and preview sentences are often used to explicitly identify topics of a text section with verbal labels.
7. **Function Identification:** Headings, preview sentences and other devices can be used to explicitly identify the function of a text section (e.g., "introduction"). This also requires verbal coding.

To illustrate these "information functions" of signaling devices, imagine that the heading for the first section of this article had been:

I. Introduction: Signaling Devices as a Challenge to TTS Software

This particular heading communicates all seven possible information functions. First, by the white space that precedes and follows it, the heading *demarcates* a boundary between distinct subsections of the text. Second, *emphasis* is applied to the heading in the form of boldface print. Third and fourth, information about both the *hierarchical* and *linear* position of the subsection is communicated by the Roman numeral "I." Fifth, the heading is a *label* for the subsection that may be used to refer uniquely to the subsection it heads. Sixth, the phrase "signaling devices as a challenge to TTS software" identifies the topic of the subsection. Finally, the label "introduction" establishes the function of the subsection. It is important to realize that not all headings serve all possible information functions. A heading usually serves only a subset of all possible functions. For example, the heading above might have been truncated to "Signaling devices as a challenge to TTS software" or simply "Introduction" or even more simply "I." All headings carry emphasis, and demarcate and label a subsection, but the other four information functions are optional where headings are concerned. Other signaling devices perform other combinations of information functions. For example: typographical contrast is often used solely to lend emphasis to specific content; preview sentences indicate a section boundary and identify a new topic; advance outlines identify topics and their hierarchical and linear organization. In short, an analysis of the information functions of various signaling devices is an important part of understanding the similarities and differences between distinguishable signaling devices.

Besides establishing the nature of similarities and differences among signaling devices, an analysis of the information functions of a given signaling device is a critical step in understanding how that signaling device might influence text-processing. The information functions of a signaling device mediate the effects of the device on text-processing. If a signal codes a particular type of information such that the information is readily accessible to cognitive processing, and if that information is relevant to the reader's goals, then that information function will influence text-processing. In a series of experiments, we have isolated each of the information functions under task conditions that make each highly relevant to task performance and shown that each information function has an influence on text processing [21, 22]. For example, if readers are asked to select sentences from a text to compose a summary, they are more likely to select a sentence from a paragraph if it is immediately preceded by a row of asterisks (i.e., demarcation) than if it is not; this is true even if the paragraph is not an appropriate source of summary information.

Of particular relevance in the present context is the observation that several information functions are often coded by purely visually means rather than verbally. It is in these cases that TTS applications often fail to render important information for the listener. Specifically, the information functions of emphasis, demarcation, hierarchical organization and linear organization are often (although not always) represented visually rather than verbally. In the heading example considered earlier, emphasis and

demarcation information was coded by typographical contrast and spacing; hierarchical and linear organization was coded by a Roman numeral in that example but it is very common for the hierarchical organization of headings to be communicated solely by formatting conventions (i.e., spacing and typographical variation). To the extent that such information is useful to the listener in identifying important content and understanding the organization of the discourse, the loss of the information can be expected to disrupt text processing. In fact, we have found that if a TTS application fails to preserve emphasis, demarcating and hierarchical information for listeners, outlining performance is severely disrupted [5].

In sum, SARA provides a very useful analysis of the distinct types of information communicated by different signaling devices. Most importantly, SARA's identification of information functions is the appropriate level of analysis for understanding how signaling devices influence on text-processing. Thus, SARA provides an answer to the important question of: What information in signaling devices must be preserved when print is converted by a TTS application?

4 Implications for Auditory Rendering of Printed Headings

SARA does not address the question: How should visually communicated information functions be represented auditorially when a printed signal is converted by a TTS application? However, it does offer two criteria that are relevant to evaluating candidate solutions to this challenge. The first is simply that the auditory signal must preserve the information functions of the printed signal; we will refer to this as the "availability criterion." Second, it must do so in a way that allows efficient cognitive processing of the information functions; we will refer to this as the "accessibility criterion." Thus, given the choice of two alternative auditory codings of the same information functions, we would select the option that more effectively and efficiently communicates those functions to the listener.

Maurel and his colleagues [34-36] have elaborated a model, the Modèle d'Oralisation des Textes Écrits pour être Lus Silencieusement or "MOTELS", that describes four possible strategies to render text visual signals in the auditory modality. Their descriptive strategy simply describes the visual properties of the text content. For instance, the sentence: "TTS systems have now reached acceptable levels of **quality** and **naturalness**" could be transformed into: "TTS systems have now reached acceptable levels of quality and naturalness. The words quality and naturalness are in bold.**"** We do not see this strategy as holding much promise because although it meets the first criterion of communicating the information function of emphasis, it does not meet the second criterion of communicating efficiently. We suspect that this strategy would be cognitively cumbersome. Further, the description of visual properties is not compatible with Design-for-All principles [37].

The second strategy in the MOTELS model is the discursive strategy. It is closely related to the descriptive strategy except that instead of describing the physical attributes of the text, the author's intentions in signaling are described. For the example above, the translation would be: "TTS systems have now reached acceptable levels of quality and naturalness. I emphasize quality and naturalness." This approach is compatible with Design-for-All principles. Although the current example would not

appear to fare well when evaluated against the cognitive accessibility criterion, there may be instantiations of this general approach that would measure up well against both the availability and accessibility criterion. For example, preceding the verbalization of a heading with a simple verbal label such as "Level 1 heading" might be an effective and efficient way to communicate demarcation, function and hierarchical information.

The third, typo-phonic strategy represents the nonverbal properties of signals by tones or earcons. Going back to the preceding example, the words "quality" and "naturalness" could be preceded by a specific sound [38]. Alternatively, the fourth, prosodic strategy attempts to find or define a prosodic equivalent to the nonverbal information communicated by a visual signal. In our example, the boldfaced words "quality" and "naturalness" might be read more slowly, more loudly and/or at a different pitch than the rest of the sentence to contrast them with the rest of the text analogous to the way that typographical contrast is used in the printed text [39-40]. This approach has been used relatively effectively in communicating emphasis in TTS applications [3], [41]. The apparent advantage of both the typo-phonic and prosodic strategies is that both should allow efficient communication. The prosodic strategy is particularly attractive in this regard in that prosodic variation is integral to the speech signal rather than requiring the addition of other information (i.e., a tone or earcon or verbal label). The possible disadvantage of each strategy may derive from the implicit nature of both categories of devices. That is, there is no natural relationship between the hierarchical status of a heading in a text and either a set of tones or earcons or variation in prosody. Further, as the number of distinctions within a particular information function increases, implicit means of communicating the distinctions may become less effective. Again taking the hierarchical status of a heading as our example, it may be easy for a listener to understand variation in tones or earcons or prosody as signaling two different hierarchical levels; however, if there are four or five levels of headings in a text, the correspondence between the auditory cues and hierarchical level may be more difficult to learn and process.

5 Conclusion

Printed signals are well-documented aids to reading and memory for expository text. However, many TTS applications fail to adequately communicate signaling and the resulting loss of information can actually disrupt text processing. SARA's analysis of printed signals might be used to augment current TTS applications by identifying what specific information should be rendered in order to preserve the signaling function. SARA also identifies two important criteria that should help guide the evaluation of alternative auditory renderings of signals. However, much work needs to be done to identify effective and efficient ways to render printed signals in auditory form.

References

1. Fellbaum, K., Kouroupetroglou, G.: Principles of electronic speech processing with applications for people with disabilities. Technology and Disability 20, 55–85 (2008)
2. Schröder, M.: Expressive speech synthesis: Past, present and possible futures. In: Tao, J., Tan, T. (eds.) Affective Information Processing, pp. 111–126. Springer, London (2009)

3. Spiliotopoulos, D., Xydas, G., Kouroupetroglou, G., Argyropoulos, V.: Experimentation on spoken format of tables in auditory user interfaces. In: Proceedings of the 11th International Conference on Human-Computer Interaction (HCII 2005), pp. 361–370 (2005)
4. Spiliotopoulos, D., Xydas, G., Kouroupetroglou, G., Argyropoulos, V., Ikospentaki, K.: Auditory universal accessibility of data tables using naturally derived prosody specification. Universal Access in the Information Society 9, 169–183 (2010)
5. Lorch Jr., R.F., Chen, H.-T., Lemarié, J.: Communicating headings and preview sentences in text and speech. Journal of Experimental Psychology: Applied 18(3), 265–276 (2012)
6. Meyer, B.J.F.: The organization of prose and its effects on memory. North-Holland, Amsterdam (1975)
7. Lorch Jr., R.F.: Text signaling devices and their effects on reading and memory processes. Educational Psychology Review 1, 209–234 (1989)
8. Lorch Jr., R.F., Lorch, E.P., Klusewitz, M.A.: Effects of typographical cues on reading and recall of text. Contemporary Educational Psychology 20, 51–64 (1995)
9. Crouse, J.H., Idstein, P.: Effects of encoding cues on prose learning. Journal of Educational Psychology 63, 309–313 (1972)
10. Fowler, R., Barker, A.: Effectiveness of highlighting for retention of text material. Journal of Applied Psychology 59, 358–364 (1974)
11. Lorch Jr., R.F., Chen, A.H.: Effects of number signals on reading and recall. Journal of Educational Psychology 78, 263–270 (1986)
12. Lorch Jr., R.F., Lorch, E.P., Matthews, P.D.: On-line processing of the topic structure of a text. Journal of Memory and Language 24, 350–362 (1985)
13. Murray, J.D., McGlone, C.: Topic overviews and processing of topic structure. Journal of Educational Psychology 89, 251–261 (1997)
14. Glynn, S.M., DiVesta, F.J.: Outline and hierarchical organization for study and retrieval. Journal of Educational Psychology 69, 69–95 (1997)
15. Lorch Jr., R.F., Lorch, E.P.: Topic structure representation and text recall. Journal of Educational Psychology 77, 137–148 (1985)
16. Lorch Jr., R.F., Lorch, E.P.: Effects of organizational signals on text processing strategies. Journal of Educational Psychology 87, 537–544 (1995)
17. Glover, J.A., Dinnel, D.L., Halpain, D.R., McKee, T.K., Corkill, A.J., Wise, S.L.: The effects of across-chapter signals on recall of text. Journal of Educational Psychology 80, 3–15 (1988)
18. Cauchard, F., Eyrolle, H., Cellier, J.-M., Hyönä, J.: Vertical perceptual span and the processing of visual signals in reading. International Journal of Psychology 45, 40–47 (2010)
19. Cauchard, F., Eyrolle, H., Cellier, J.-M., Hyönä, J.: Visual signals vertically extend the perceptual span in searching a text: A gaze-contingent window study. Discourse Processes 47, 617–640 (2010)
20. Klusewitz, M.A., Lorch Jr., R.F.: Effects of headings and familiarity with a text on strategies for searching a text. Memory & Cognition 28, 667–676 (2000)
21. Lorch Jr., R.F., Lemarié, J., Grant, R.A.: Signaling hierarchical and sequential organization in expository text. Scientific Studies of Reading 15, 267–284 (2011)
22. Lorch Jr., R.F., Lemarié, J., Grant, R.A.: Three information functions of headings: A test of the SARA theory of signaling. Discourse Processes 48, 139–160 (2011)
23. Hyönä, J., Lorch Jr., R.F.: Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. Learning and Instruction 14, 131–152 (2004)

24. Hyönä, J., Lorch Jr., R.F., Kaakinen, J.: Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. Journal of Educational Psychology 94, 44–55 (2002)
25. Sanchez, R.P., Lorch, E.P., Lorch Jr., R.F.: Effects of headings on text processing strategies. Contemporary Educational Psychology 26, 418–428 (2001)
26. Lorch Jr., R.F., Lorch, E.P.: Effects of organizational signals on free recall of expository text. Journal of Educational Psychology 88, 38–48 (1996)
27. Mayer, R.E., Dyck, J.L., Cook, L.K.: Techniques that help readers build mental models from scientific text: Definitions pretraining and signaling. Journal of Educational Psychology 76, 1089–1105 (1984)
28. Lorch Jr., R.F., Lorch, E.P., Inman, W.E.: Effects of signaling topic structure on text recall. Journal of Educational Psychology 85, 281–290 (1993)
29. Surber, J.R., Schroeder, M.: Effect of prior domain knowledge and headings on processing of informative text. Contemporary Educational Psychology 32, 485–498 (2007)
30. Dee-Lucas, D., Larkin, J.H.: Novice strategies for processing scientific texts. Discourse Processes 9, 329–354 (1986)
31. Dee-Lucas, D., Larkin, J.H.: Novice rules for assessing importance in science texts. Journal of Memory and Language 27, 288–308 (1988)
32. Meyer, B.J.F., Brandt, D.M., Bluth, G.J.: Use of top-level structure in text: Key for reading comprehension of ninth grade students. Reading Research Quarterly 16, 72–103 (1980)
33. Lemarié, J., Lorch, R.F., Eyrolle, H., Virbel, J.: SARA: A Text-Based and Reader-Based Theory of Text Signaling. Educational Psychologist 43(1), 27–48 (2008)
34. Lemarié, J., Eyrolle, H., Cellier, J.-M.: Visual signals in text comprehension: How to restore them when oralizing a text via speech synthesis. Computers in Human Behavior 22, 1096–1115 (2006)
35. Maurel, F.: De l'écrit à l'oral: analyses et générations. In: Actes de la 11ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2004), Fès, Maroc, pp. 289–298 (Avril 2004)
36. Maurel, F., Lemarié, J., Vigouroux, N.: Oralisation de structures visuelles: de la lexico-syntaxe à la prosodie. In: Actes du colloque national Interface Prosodique (IP 2003), Nantes, France, Mars 27-29, pp. 137–142 (2003)
37. Tsonos, D., Xydas, G., Kouroupetroglou, G.: Auditory accessibility of metadata in books: A design for all approach. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 436–445. Springer, Heidelberg (2007)
38. McGookin, D.K., Brewster, S.A.: Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. ACM Transactions on Applied Perceptions 1, 130–155 (2004)
39. Giraud, S., Thérouanne, P.: Role of lexico-syntactic and prosodic cues in spoken comprehension of enumeration in sighted and blind adults. Presented at Multidisciplinary Approaches to Discourse, Moissac, France (2010)
40. Giraud, S., Uzan, G., Thérouanne, P.: L'accessibilité des interfaces informatiques pour les déficients visuels. In: Dinet, J., Bastien, C. (eds.) L'ergonomie Des Objets et Environnements Physiques et Numériques, Hermes - Sciences Lavoisier, Paris (2011)
41. Argyropoulos, V.S., Sideridis, G.D., Kouroupetroglou, G., Xydas, G.: Auditory discriminations of typographic attributes of documents by students with blindness. British Journal of Visual Impairment 27, 183–203 (2009)