LEGIBILITY OF PERCEPTUALLY-TUNED GRAYSCALE FONTS

Kevin O'Regan, Nicole Bismuth, Laboratoire de Psychologie Expérimentale, CNRS, Paris
Roger D. Hersch, Alexandros Pappas, Ecole Polytechnique Fédérale, Lausanne
oregan@ext.jussieu.fr, hersch@di.epfl.ch

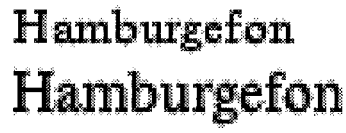
ABSTRACT

Perceptually-tuned grayscale fonts are generated from character outline descriptions by applying to them a set of modifications specifically conceived for strengthening thin character parts, obtaining well-contrasted bars and preserving important relationships between character shape parts. The present study aims at comparing the legibility of perceptually-tuned grayscale and bilevel display fonts at small and very small sizes (6, 8 and 10 pt). The study confirms the results of previous studies indicating that reading speed is to a large extent independent of the typography (bilevel or grayscale) and the font size. However, perceptually-tuned grayscale characters perform better than bilevel characters for an italic string search task in a meaningless text. Regarding the subjective preferences of the test subjects, perceptually-tuned grayscale fonts at 8 and 10 point sizes received a superior rating than bilevel fonts at the same sizes.

1. PERCEPTUALLY TUNED GRAYSCALE FONTS

In order to improve the display of text on limited resolution computer displays, researchers have tried to trade-off the lack of spatial resolution for an increased number of intensity levels [1, 2, 3]. Especially with respect to typographic text display on resolution limited CRT and LCD display devices, it has been shown that using grayscale makes font dependent character features visible which disappear when displaying text with bilevel characters. Though they can more accurately render font differences, grayscale characters generated by filtering and resampling high-resolution bilevel master characters nevertheless look rather fuzzy.

a) Box filter



b) Manually edited

Figure 1. Characters generated by filtering and resampling compared with manually edited grayscale characters

Different instances of similar character structure elements such as vertical bars or curved stems do not have similar intensity profiles (Fig. 1a). This is due to the fact that the original bilevel master character incorporates frequencies well beyond the Nyquist limit (1/2 the resampling frequency) and that LCD displays are capable of displaying individual pixels as constant intensity squares of unit size, thereby enabling high intensity gradients between neighbouring pixels.

The phase of character elements with respect to the target grid influences their intensity profiles (Fig. 2a).

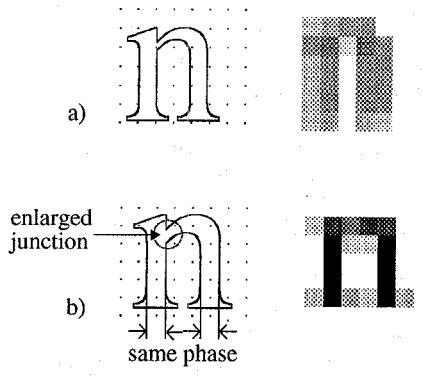


Figure 2. The phase of the vertical bars defines their intensity profile: (a) without outline modifications and (b) with outline modifications

In order to automatically generate improved quality grayscale characters similar to the characters which would be designed manually by skilled typographic designers, we have created a software package which applies modifications to the outline of the characters according to information specifying character structure elements such as bars, curved stems and serifs [4]. These modifications ensure that the left edge of vertical bars have a sharp contrast and that similar bars and curved stems have similar intensity profiles [5]. Furthermore, we preserve the structure of characters by enlarging character parts (Fig. 2b) which would be too thin to be displayed correctly.

2. COMPARING THE LEGIBILITY OF PERCEPTUALLY-TUNED AND BILEVEL FONTS

At a first glance, the perceptually-tuned grayscale characters seem to be more readable and to offer improved reading comfort, especially at small sizes (Fig. 3).

The authors therefore decided to conduct a legibility study and

0-7803-3258-X/96/$5.00 © 1996 IEEE

to compare the perceptually-tuned grayscale and bilevel typographies for small character sizes.

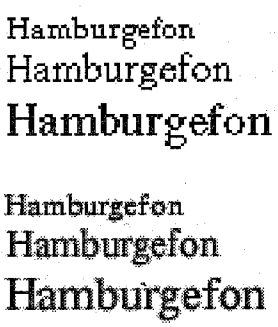


Figure 3. Automatically generated bilevel and perceptually-tuned grayscale characters

A previous study comparing the legibility of traditional grayscale characters with that of bilevel characters showed that there was no significant difference in reading speeds between the two typographies at 10 point size [6], but that font identification was much easier with grayscale fonts than with bilevel fonts. Regarding esthetic judgement, there was no preference for the grayscale rather than the bilevel typography.

In order to compare perceptually-tuned grayscale fonts with bilevel ones, we start from the hypothesis that the impact of perceptually tuned grayscale fonts should be the greatest at small sizes, where bilevel fonts are difficult to read. We therefore decided to conduct the legibility tests with font sizes 10, 8 and 6 points, with corresponding interline spaces of 10/72", 8/72" and 6/72". We defined a reading task where the subjects read text displayed with bilevel and with perceptually-tuned grayscale fonts at the chosen font sizes of 10, 8 and 6 points.

In order to compare the impact of the typographies in an additional task involving visual search, the subjects were asked to indicate as fast as fast as possible the first three italic strings they found in a text composed of meaningless character strings. This meaningless text was generated from a normal text with the characters scrambled to create meaningless words.

Finally, the subjects were asked to show their preferences by giving an evaluation mark in the range 1 - 10 to each of the 6 combinations of typographies and point sizes.

The tests were conducted on a Macintosh Powerbook 180c, with an 8.5" LCD display having a usable display surface of 640 pixels horizontally and 448 pixels vertically, at a display resolution of 95.5 dpi. Characters were displayed as black or gray on a constant intensity white background. Each of the 12 subjects was asked to perform the reading task, the search task and to give his or her preferences. Reading times and search times were measured in milliseconds. Subjects were 25 to 39 years old, with normal or corrected vision, and had no previous experience in typography. The experiment was carried out at the Laboratory of Experimental Psychology of University Paris VI.

Each subject was asked to read six different text pages, each page representing a different experiment, i.e. font sizes 6, 8 and 10 point and bilevel or respectively perceptually tuned grayscale typography (Helvetica characters). The sequence in which the texts were read was different for each of the subjects. Secondary factors such as ambiant light, display luminance, display contrast and reading distance were kept constant.

3. COMPARATIVE LEGIBILITY RESULTS

Figure 4 shows the mean reading times and their associated 95% confidence intervals for the chosen test sizes and typographies. Mean reading times are nearly the same for the bilevel and for the perceptually tuned grayscale typography. They are also fairly constant throughout the different font sizes. This result is confirmed by previous research, which has shown that reading time is not strongly influenced by character size, at least at font sizes larger than 8 point [7]. Similar results have been obtained more recently by several researchers [8], [9]. Nevertheless, it is remarkable that even at 6 point, reading speed remains constant. It seems that the reading process relies heavily on background knowledge to identify partly degraded words [10].

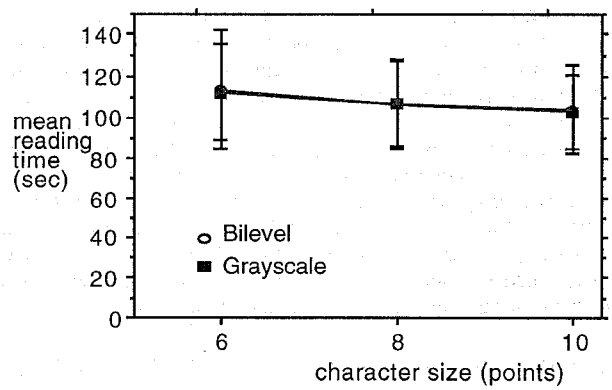


Figure 4. Mean reading times and associated 95% confidence intervals for reading text for character sizes 6, 8, 10 points and for bilevel and grayscale typographies.

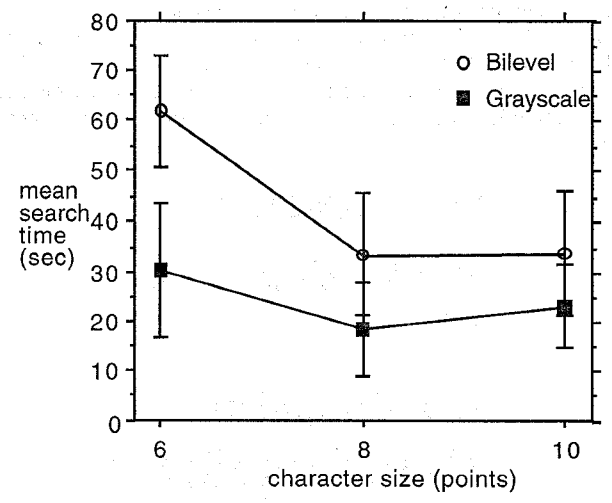


Figure 5. Mean search times and associated 95% confidence intervals for the italic string searching task

Figure 5 shows the mean search times and their 95% confidence intervals for the italic string searching task. Figure 6 shows a portion of the meaningless text with the italic string, at size 6, displayed with perceptually tuned grayscale and bilevel typographies. At all sizes, the search task is executed in significantly

less time with perceptually-tuned grayscale fonts than with bilevel fonts. This difference is even more pronounced at very low font sizes. Analysis of variance confirms the significance of the font size and of the typography (grayscale versus bilevel). This result shows that for the italic search task the physical aspect of letterforms is of primary importance, in contrast to the pure reading task, where existing knowledge and cognitive processing are of primary importance.

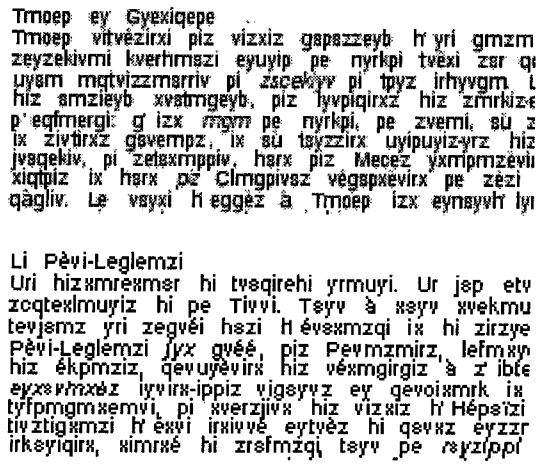


Figure 6. Extract of meaningless text with italic strings used for the search task, displayed with perceptually tuned grayscale and bilevel typographies

Figure 7 shows in terms of mean values the preferences of the subjects regarding the typographies and the point sizes and their 95% confidence intervals for evaluation marks ranging between 1 and 10. There is a clear preference for perceptually-tuned grayscale characters at point sizes 8 and 10. Regarding font size 6, there was no preference for the perceptually-tuned grayscale typography, even though at this size the search task was more successful with perceptually-tuned rather than with bilevel fonts. As shown in figure 8, at point size 6, the character width varies too much between the different characters and makes this instance of perceptually tuned grayscale font less acceptable than at sizes 8 and 10, where characters have a coherent and uniform appearance. This may explain the lack of preference for perceptually tuned grayscale characters at point size 6.

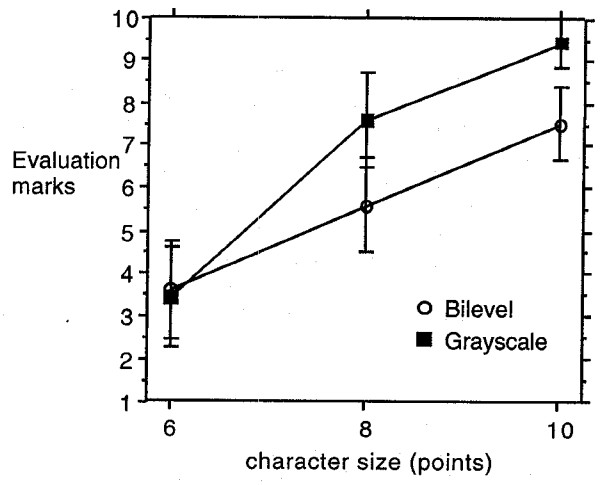


Figure 7. Mean preference values as a function of typographies and point sizes and associated 95% confidence intervals

Since the study by Black & Boag [6] revealed no preference for traditional grayscale characters at 10 points, the preference for perceptually-tuned grayscale characters at point sizes 8 and 10 clearly distinguishes the perceptually-tuned grayscale character generation method from previous grayscaling methods.

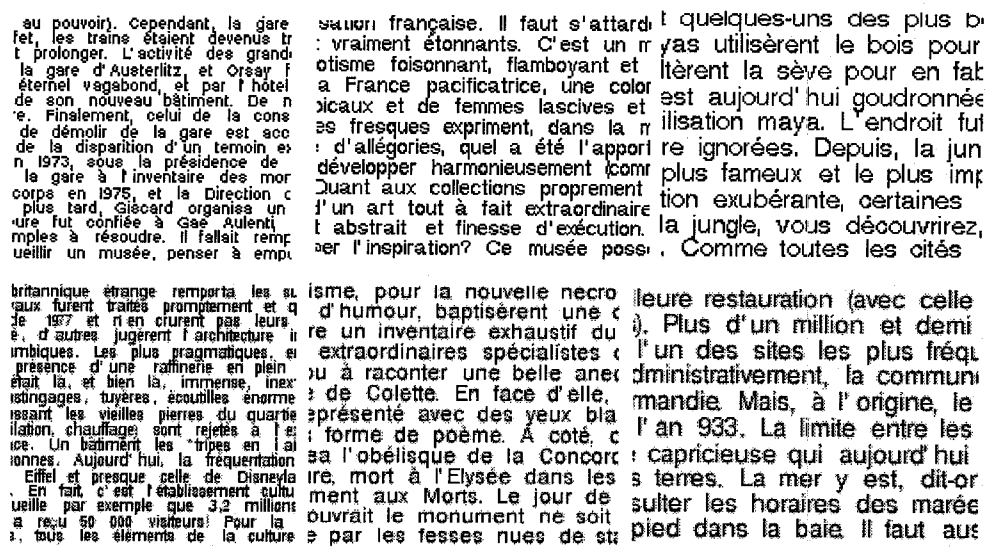


Figure 8. Portions of text used for giving evaluation marks with respect to typographies and point sizes (left 6 pt, middle 8 pt, right 10pt, top bilevel typography, bottom perceptually-tuned grayscale)

4. CONCLUSIONS

The deficiencies of traditional character grayscaling methods are briefly discussed and the basics of the new perceptually-tuned grayscale character generation method are described. A study was conducted to compare the legibility of perceptually-tuned grayscale and bilevel display fonts at small and very small sizes. This study confirmed the results of previous studies indicating that reading speed is to a large extent independent of the typography (bilevel or grayscale) and the font size. Perceptually-tuned grayscale characters performed much better than bilevel characters for an italic string search task in a meaningless text. This search task is highly sensitive to differences in typographies and could be used as a testing tool for further typographic evaluation tasks. Regarding the subjective preferences of the test subjects, perceptually-tuned grayscale fonts at 8 and 10 point sizes were given a clearly superior rating compared to bilevel fonts at these sizes. This result is to be related to the study conducted by Black and Boag [6], where there was no user preference for traditional grayscale fonts versus bilevel fonts at 10 point size. This difference in user preference may be attributed to the improved visual quality of perceptually-tuned grayscale fonts.

The present study did not take into account the density of text on the display surface. Bilevel and grayscale typographies were generated automatically from outline fonts without introducing any line length constraints. The comparative examples show that perceptually-tuned grayscale fonts are more compact than bilevel fonts. With the introduction of line length constraints, the bilevel characters would need to be spaced more tightly and characters would partly interfere with one another, as is the case in display editors, where the displayed text is positioned exactly as on the printed page. We therefore intend to repeat the same tests as in the present study, but this time introducing line length or text density constraints.

5. REFERENCES

[1] F. Crow, "The Use of Grayscale for Improved Raster Display of Vectors and Characters", Computer Graphics (Proc. Siggraph), 12 (3), 1–6, 1978.

[2] J. Warnock, "The Display of Characters Using Gray Level Sample Arrays", Computer Graphics (Proc. Siggraph), 14 (3), pp. 302–307, 1980.

[3] A. Naiman, A. Fournier, "Rectangular Convolution for Fast Filtering of Characters", Computer Graphics (Proc. Siggraph), 21 (4), pp. 233–242, 1987.

[4] R. D. Hersch, "Font Rasterization, the State of the Art", *Visual and Technical Aspects of Type*, Cambridge University Press, pp. 78–109, 1993.

[5] R. D. Hersch, C. Bétrisey, J. Bur, A. Gürtler, "Perceptually-Tuned Generation of Grayscale Fonts", IEEE Computer Graphics and Applications, 15 (6), pp. 78–89, 1995.

[6] A. Black, A. Boag, "Choosing Binary or Greyscale Bitmaps: some Consequences for Users". In C. Vanoirbeek & G. Coray (Eds.) *Proceedings Electronic Publishing*, Cambridge University Press, pp. 247–260, 1992.

[7] E. Javal, *Physiologie de la lecture et l'écriture*, Paris, Alcan, 1905.

[8] J. D. Gould, J. D., L. Alfaro , R. Finn, B. Haupt, A. Minuto, "Reading from CRT displays can be as fast as reading from paper". *Human factors*, 29, pp. 497–51, 1987.

[9] G. Legge, G.S. Rubin, A. Luebker, "Psychophysics of reading - V. The role of contrast in normal vision", *Vision Research*, 27, pp. 1165–1177, 1987.

[10] J. McClelland, D.E. Rumelhart, "An interactive activation model of context effects in letter perception : Part 1. An account of basic findings", *Psychological review*, 88, pp. 375–408, 1981.