Acoustic Mapping of Visual Text Signals through Advanced Text-to-Speech: the Case of Font Size

GEORGIOS KOUROUPETROGLOU
Division Communications and Signal Processing
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Panepistimioupolis, Ilissia, GR-15784 Athens
GREECE
koupe@di.uoa.gr

Abstract: - Current Text-to-Speech systems, commonly used for document accessibility, do not include an effective and standard acoustic provision of the visual typographic cues embedded in them. In this work, we first introduce the text signals (i.e. the writing devices that emphasize aspects of a text's content or structure) along with an appropriate architecture for the structure of documents as well as the main principles and technological topics of document accessibility. Then, the emerging technological approach of Document-to-Audio (DtA) we have developed is presented. DtA essentially constitutes the next generation of Text-to-Speech systems that supports the acoustic mapping of visual text-signals and thus provides much better accessibility to documents. Finally, for the case of the font size in the typographic layer, we present the results of two quantitative approaches (direct mapping and emotional-based mapping) for the rendering of text signals through DtA.

Key-Words: - Text-to-Speech, Text Signals, Document-to-Audio, Document Visual Elements, Accessibility

1 Introduction

Document accessibility enables an electronic document, either in the form of a common file format (such as .pdf .doc, .ppt) or a Web document, to be used effectively, efficiently and satisfactory by more users in more situations or context of use. It concerns all the aspects of document functionality that includes browsing, searching, navigation and reading [1]. Basic accessibility of documents is provided by the Text-to-Speech systems [2, 47, 48].

Accessibility of documents is very important not only for the print-disabled readers, i.e. those with vision impairment (blindness, low vision, color blindness, dyschromatopsia, etc.), a learning disability (including dyslexia) or a motor disability (such as loss of dexterity that prevents the physical handling of a document), but also for those with an occasional or situational "disability". A typical example is a user who wants to access a document while driving a car, i.e. his/her eyes and hands are busy.

In this paper, we present our recent efforts towards advanced document accessibility by supporting not only the accessibility of the text content of a document, but also the accessibility of the text signals, i.e. the writing devices that emphasizes aspects of a text's content or structure. After the introduction of an appropriate document's architecture, we present the main principles and technological topics of document accessibility. Then, the emerging technological approach of Document-to-Audio (DtA) we have developed is presented for the universal rendering of text signals to auditory modality. DtA is based on the principles of Design for All. DtA essentially constitutes the next generation of Text-to-Speech systems and they are capable to provide much better accessibility to documents. Finally, for the case of text signals derived from the font size in the typographic layer, we present the results of two quantitative approaches (direct mapping and emotional-based mapping) for the rendering of text signals through DtA.

2 Text Signals in Documents

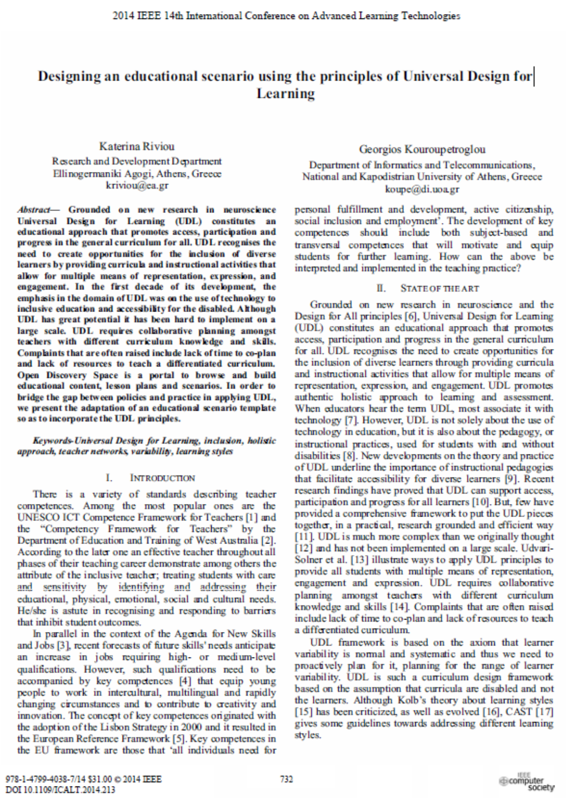
2.1 Document Architecture

According to the communication theorist Marshall McLuhan [3], a document is the "medium" in which a "message" (information) is communicated. The content of the printed or electronic documents includes mainly the text and the images (i.e. figures, drawings, graphs, pictures, charts, diagrams, schematic illustrations, maps, photos, etc.). Furthermore, it may include mathematical or in general scientific symbols and formulas. The term text document refers to the textual content only of a document.

Besides its content, a printed or electronic document contains a number of presentation elements or attributes that apply on its text content (Figure 1): a) design glyphs or typographic elements (i.e. visual representation of letters and characters in a specific font and style) and b) arrangement of the content on the page and the document as a whole. For example, the title of a chapter can be recognized by placing it at the top of the page and in larger font size than the body of the text. Moreover, text color, but also the bold font style, can be used to indicate emphasis in a specific part of a text document.

Figure 1: Presentation elements of a document

Rich-text content (Figure 2) is a text document that preserves all its presentation elements. In contrast, a plain-text document ignores the presentation elements. Most of the current Text-to-Speech (TtS) systems essentially use documents in plain text [19, 46].



2014 IEEE 14th International Conference on Advanced Learning Technologies

Designing an educational scenario using the principles of Universal Design for Learning

Katerina Riviou
Research and Development Department
Ellinogermaniki Agogi, Athens, Greece
kriviou@ea.gr

Georgios Kouroupetroglou
Department of Informatics and Telecommunications,
National and Kapodistrian University of Athens, Greece
koupe@di.uoa.gr

***Abstract*— Grounded on new research in neuroscience Universal Design for Learning (UDL) constitutes an educational approach that promotes access, participation and progress in the general curriculum for all. UDL recognises the need to create opportunities for the inclusion of diverse learners by providing curricula and instructional activities that allow for multiple means of representation, expression, and engagement. In the first decade of its development, the emphasis in the domain of UDL was on the use of technology to inclusive education and accessibility for the disabled. Although UDL has great potential it has been hard to implement on a large scale. UDL requires collaborative planning amongst teachers with different curriculum knowledge and skills. Complaints that are often raised include lack of time to co-plan and lack of resources to teach a differentiated curriculum. Open Discovery Space is a portal to browse and build educational content, lesson plans and scenarios. In order to bridge the gap between policies and practice in applying UDL, we present the adaptation of an educational scenario template so as to incorporate the UDL principles.**

Keywords-Universal Design for Learning, inclusion, holistic approach, teacher networks, variability, learning styles

I. INTRODUCTION

There is a variety of standards describing teacher competences. Among the most popular ones are the UNESCO ICT Competence Framework for Teachers [1] and the "Competency Framework for Teachers" by the Department of Education and Training of West Australia [2]. According to the later one an effective teacher throughout all phases of their teaching career demonstrate among others the attribute of the inclusive teacher; treating students with care and sensitivity by identifying and addressing their educational, physical, emotional, social and cultural needs. He/she is astute in recognising and responding to barriers that inhibit student outcomes.

In parallel in the context of the Agenda for New Skills and Jobs [3], recent forecasts of future skills' needs anticipate an increase in jobs requiring high- or medium-level qualifications. However, such qualifications need to be accompanied by key competences [4] that equip young people to work in intercultural, multilingual and rapidly changing circumstances and to contribute to creativity and innovation. The concept of key competences originated with the adoption of the Lisbon Strategy in 2000 and it resulted in the European Reference Framework [5]. Key competences in the EU framework are those that 'all individuals need for personal fulfillment and development, active citizenship, social inclusion and employment'. The development of key competences should include both subject-based and transversal competences that will motivate and equip students for further learning. How can the above be interpreted and implemented in the teaching practice?

II. STATE OF THE ART

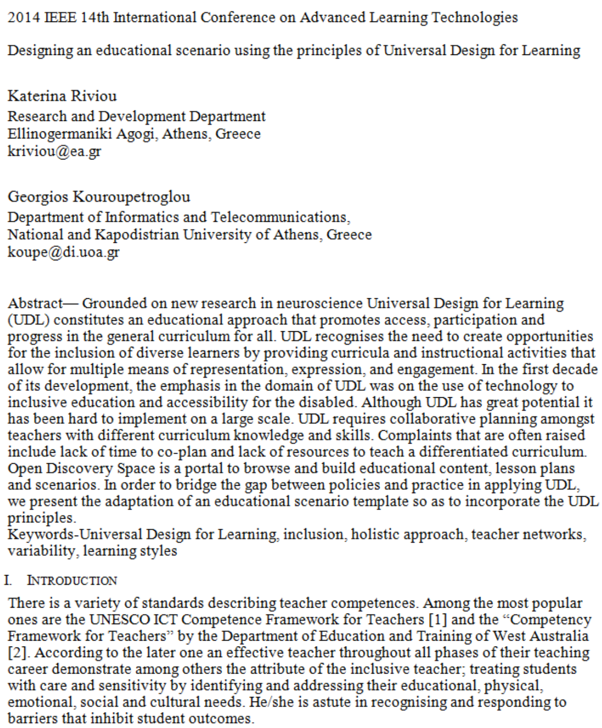
Grounded on new research in neuroscience and the Design for All principles [6], Universal Design for Learning (UDL) constitutes an educational approach that promotes access, participation and progress in the general curriculum for all. UDL recognises the need to create opportunities for the inclusion of diverse learners through providing curricula and instructional activities that allow for multiple means of representation, expression, and engagement. UDL promotes authentic holistic approach to learning and assessment. When educators hear the term UDL, most associate it with technology [7]. However, UDL is not solely about the use of technology in education, but it is also about the pedagogy, or instructional practices, used for students with and without disabilities [8]. New developments on the theory and practice of UDL underline the importance of instructional pedagogies that facilitate accessibility for diverse learners [9]. Recent research findings have proved that UDL can support access, participation and progress for all learners [10]. But, few have provided a comprehensive framework to put the UDL pieces together, in a practical, research grounded and efficient way [11]. UDL is much more complex than we originally thought [12] and has not been implemented on a large scale. Udvari-Solner et al. [13] illustrate ways to apply UDL principles to provide all students with multiple means of representation, engagement and expression. UDL requires collaborative planning amongst teachers with different curriculum knowledge and skills [14]. Complaints that are often raised include lack of time to co-plan and lack of resources to teach a differentiated curriculum.

UDL framework is based on the axiom that learner variability is normal and systematic and thus we need to proactively plan for it, planning for the range of learner variability. UDL is such a curriculum design framework based on the assumption that curricula are disabled and not the learners. Although Kolb's theory about learning styles [15] has been criticized, as well as evolved [16], CAST [17] gives some guidelines towards addressing different learning styles.

978-1-4799-4038-7/14 $31.00 © 2014 IEEE
DOI 10.1109/ICALT.2014.213

732

(a)



2014 IEEE 14th International Conference on Advanced Learning Technologies

Designing an educational scenario using the principles of Universal Design for Learning

Katerina Riviou
Research and Development Department
Ellinogermaniki Agogi, Athens, Greece
kriviou@ea.gr

Georgios Kouroupetroglou
Department of Informatics and Telecommunications,
National and Kapodistrian University of Athens, Greece
koupe@di.uoa.gr

Abstract— Grounded on new research in neuroscience Universal Design for Learning (UDL) constitutes an educational approach that promotes access, participation and progress in the general curriculum for all. UDL recognises the need to create opportunities for the inclusion of diverse learners by providing curricula and instructional activities that allow for multiple means of representation, expression, and engagement. In the first decade of its development, the emphasis in the domain of UDL was on the use of technology to inclusive education and accessibility for the disabled. Although UDL has great potential it has been hard to implement on a large scale. UDL requires collaborative planning amongst teachers with different curriculum knowledge and skills. Complaints that are often raised include lack of time to co-plan and lack of resources to teach a differentiated curriculum. Open Discovery Space is a portal to browse and build educational content, lesson plans and scenarios. In order to bridge the gap between policies and practice in applying UDL, we present the adaptation of an educational scenario template so as to incorporate the UDL principles.
Keywords-Universal Design for Learning, inclusion, holistic approach, teacher networks, variability, learning styles

I. INTRODUCTION

There is a variety of standards describing teacher competences. Among the most popular ones are the UNESCO ICT Competence Framework for Teachers [1] and the "Competency Framework for Teachers" by the Department of Education and Training of West Australia [2]. According to the later one an effective teacher throughout all phases of their teaching career demonstrate among others the attribute of the inclusive teacher; treating students with care and sensitivity by identifying and addressing their educational, physical, emotional, social and cultural needs. He/she is astute in recognising and responding to barriers that inhibit student outcomes.

(b)

Figure 2: Rich-text document (a) and plain-text document (b) without formatting

The elements of a text document can be classified in three layers (Figure 3) [4]:

- Logical layer: it associates content with structural elements such as headings, titles/ subtitles, chapters, paragraphs, tables, lists, footnotes, and appendices.
- Layout layer: it associates content with architectural elements relating to the arrangement on pages and areas within pages, such as margins, columns, alignment and orientation (portrait or landscape).
- Typography layer: it includes font (type, size, color, background color, etc.) and font style such as bold, italics, underline.

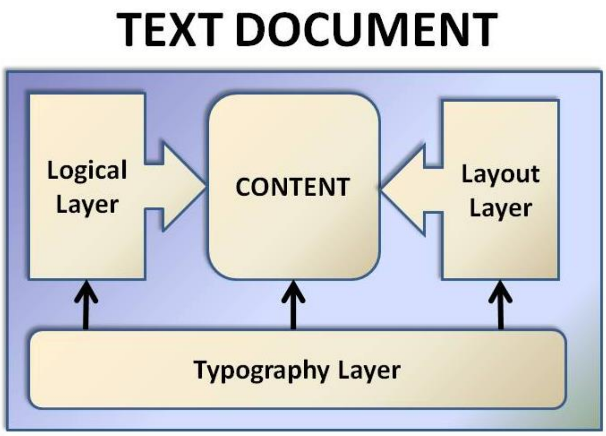


Figure 3: Text document architecture

The above three layers are complementary and not independent. Typography can be applied to both the logical and the layout layers of a document. Moreover, typography can be applied to the main body of the text directly. For example, a word in bold can be used either for the introduction of a new term or to indicate a person's name. Also, a heading can be arranged in the center of a line (layout layer).

The organization of a document can be classified into two main aspects: the logical and the physical. The logical layer of the document architecture defined above corresponds to its logical organization with the same elements (e.g. headings, titles/ subtitles, chapters, paragraphs, tables, lists, footnotes, and appendices). At the page level, the physical organization of a document is described by its layout layer in connection with the physical realization of a number of logical layout elements (e.g. headings, titles/ subtitles, paragraphs, tables, lists, footnotes). The organization of a printed of electronic multipage document as a whole corresponds with the physical implementation of a part of its logical layer elements (e.g. chapters, appendices, indexed, references). The organization of a document is domain specific (e.g. text book, scientific paper, technical report, newspaper, magazine).

2.2 Text Signals

The term text signal has been proposed [5] as the writing device that emphasizes aspects of a text's content or structure carrying semantic information over and above the content. It attempts to pre-announce or emphasize content and/or reveal content relationship [5-7, 49-51]. Headings or titles in text documents are considered as signals [6]. Also, "*input enhancement*" is an operation whereby the saliency of linguistic features is augmented through e.g. textual enhancement for visual input (i.e. bold) and phonological manipulations for aural input (i.e. oral repetition) [52].

All the text signalling devices, either mentioned as text signals or text layers: a) share the goal for directing the reader's attention during reading, b) facilitate specific cognitive process occurring during reading, c) ultimate comprehension of text information, d) may influence memory on text and e) direct selective access between and within texts [5].

On the other hand, it seems that there is a plethora of semantics in applying the typographic layer. For example, in contrast to the tags introduced by the W3C for the bold and italics font styles [32], we have identified [41] the following eight different "labels" that the readers seem to use most frequently in order to semantically characterize text in "bold" and "italics (a total of 2.927 entities, of which 1.866 were occurrences of "bold" and 1.061 of "italics" were manually labelled in a corpus of 2.000 articles of a Greek e-newspaper):

- **Emphasis**: A word or phrase that is considered significant and needs to be stressed out.
- **Important / Salient**: A word or phrase, which is near or is part of a piece of information that is considered important and should be noticeable. A word or phrase that "catches the eye" of the reader.
- **Basic Block**: A block of text, which introduces or summarizes the main content of the article.
- **Quotation**: A piece of text corresponding to a fragment of written or oral expression of a person other than the writer of the article.
- **Note**: A piece of text serving at providing additional information or explanation related to a part or to the whole of the article.
- **Title**: A piece of text corresponding to the name of a movie, play, book and so on or the

title of a newspaper, television channel or journal.
- **List / Numeration Category**: A word of phrase that is part of a list or a numeration and serves as a new “instance” indicator.
- **Interview / Dialogue**: A piece of text that is part of an interview (the question or the answer) or that corresponds to a dialogue between two persons.

A few “logical” labels, such as “subtitle” or “footnote” were also mentioned by some readers, but these were not considered to be “semantic” labels and were therefore not of value for the purposes of this study.

3 Document Accessibility

The accessibility of documents can be analyzed in the following four dimensions: technological, legal, economic and social/ethical.

The Convention on the Rights of Persons with Disabilities (CRPD) [8] of the United Nations, adopted on 13 Dec. 2006 by the General Assembly and effective from 3 May 2008, and specifically Article 9, is the main international legal instrument for document accessibility. Up the end of 2012, 126 countries ratified and adopted CRPD on a national legislation level. The Americans with Disabilities Act [9] enacted by the U.S. Congress in 1990 as well as the Federal procurement of electronic and information technology [10] cover the issue of document accessibility. i2010 [11], the European Union’s respective policy framework for the information society and media 2005-2009, along with the Digital Agenda for Europe [12], incorporate the Commission’s strategy involving with document accessibility.

The economic aspects of document accessibility include on one hand the development cost of accessible documents and on the other the return from their production and distribution (e.g. increase the range of possible customers and/or decrease the production of alternative formats). By applying the Design-for-All (or universal design) principle, the extra cost one has to spend for a document to become accessible can be sufficiently small. Universal access embraces theoretical, methodological, and empirical research of both a technological and non-technological nature that addresses accessibility, usability, and, ultimately, acceptability of information society technologies by anyone, anywhere, at any time, and through any media and device [45]. In the framework of universal access [13], we define universal design or Design-for-All as the general framework catering for conscious and systematic efforts to proactively apply principles, methods, and tools to develop documents that are accessible and usable by more people, including the youngest, the elderly and people with different types of disabilities, in more situations or context of use, thus avoiding the need for a posteriori adaptations, or specialized design.

Social/ethical issues of document accessibility promote equal opportunities, without exclusion and are part of the Corporate Social Responsibility of an institution, either public or private.

The basic technological requirements for a document to be accessible are:
- Content must be perceivable: information must be presentable to users in ways (such as other modalities) they can perceive, i.e. content cannot be unobservable to all of their senses.
- The reading platform/agent (hardware and/or software) must be operable concerning user interface components as well as navigation, browsing and searching, i.e. the interface cannot require interaction that a user cannot perform.
- Content must be understandable: information and the operation of user interface must be understandable, i.e. the content or operation cannot be beyond the understanding of the users.
- Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies, i.e. as technologies and user agents evolve, the content should remain accessible.

Accessible documents are related to alternate media or alternative formats, including: e-document (accessible files in MS-Word, PDF, Open Office, LaTEX), large print, braille, tactile graphics, audio files, DAISY and accessible Web content. Most online e-documents, audio books or e-book readers or are not accessible. Any electronic document (such as a file in MS-Word, PDF or Open Office) is not accessible per se. A number of guidelines have been introduced for developing accessible electronic documents [14]. Microsoft provides detailed guidelines to create accessible documents, spreadsheets, and presentations with Microsoft Office [15] and has included accessibility checker functionality. Adobe provides a series of accessibility guides to assist authors to create PDF files that are accessible to people with disabilities [16]. Moreover, MS Word and PowerPoint plug-ins exist that help to create accessible PDF documents [17].

Accessibility essentially ensures compatibility between the content of a document and the technology/agent of the reader. Within the scope of this work, a user agent is any software that retrieves, renders and facilitates end user interaction with the content of a document. An electronic document can be read by mainstream platforms, such as personal computers (either desktops or laptops), as well as by mobile devices (smart-phones and tablets). Assistive Technology (AT) includes products, devices, equipment or software applications that are used to maintain, increase or improve the functional capabilities of individuals with disabilities. AT is designed to enable independence for disabled and older people. The Human Activity Assistive Technology (HAAT) model introduced by Cook & Hussey [18] describes the overall goal of an AT system. There is a variety of AT for document accessibility, ranging from low-tech devices (e.g. portable magnifiers) to advanced Automated Reading Devices [19].There are clear benefits of using AT [20] for the activities and the participation of persons with disabilities according to the International Classification of Functioning, Disability and Health [21]. Typical AT to access documents include:

- Screen readers [19],
- Text-to-Speech [2, 19]
- Screen Magnifiers [22],
- CCTV (Closed Circuit TeleVision) systems [23, 24],
- Alternative input devices (e.g. switches, joysticks, eye gazing) [25, 26],
- Braille displays [27],
- Automated Reading Devices (ARD) [19],
- DAISY readers [28]: software applications or portable hardware devices for playing e-books that follow the ANSI/NISO Z39.86-2005 specifications for the Digital Talking Book.

The Design-for-All principle has been adopted by the main computer operating systems (MS-Windows, Linux, Mac OS, Android) and office applications (MS-Office, OpenOffice, Adobe PDF).

The Web Accessibility Initiative (WAI) [31] of the World Wide Web Consortium [32] promotes the implementation of accessibility improvements in Web technologies through the development of a set of three WAI guidelines as Recommendations: Web Content Accessibility Guidelines (WCAG), User Agent Accessibility Guidelines (UAAG), and Authoring Tool Accessibility Guidelines (ATAG). These guidelines describe features needed to achieve different levels of accessibility, and include reference checklists and implementation techniques. It is essential that several different components of Web development and interaction work together in order for the Web to be accessible to people with disabilities. These components include [31]:

- content - the information in a Web page or Web application, including:
  - natural information, such as text, images, and sounds,
  - code or markup that defines structure, presentation, etc.,
- Web browsers, media players, and other "user agents"
- Assistive Technology, in some cases - screen readers, alternative keyboards, switches, scanning software, etc.
- users' knowledge, experiences, and in some cases, adaptive strategies using the Web
- developers - designers, coders, authors, etc., including developers with disabilities and users who contribute content
- authoring tools - software that creates Web sites
- evaluation tools - Web accessibility evaluation tools, HTML validators, CSS validators, etc.

Web developers usually use authoring tools and evaluation tools to create Web content. People ("users") use Web browsers, media players, Assistive Technologies, or other "user agents" to get and interact with the content.

The WAI Technical Activity develops techniques to improve tools for evaluation and repair of Web sites. Web Accessibility Guidelines introduced by the Web Accessibility Initiative [31] of the World Wide Web Consortium [32] play a very significant role in the domain of web content accessibility. These guidelines concern all the Web content developers (page authors and site designers) as well as the developers of authoring tools. Following these guidelines the Web documents become a device / software independent. Web Content Accessibility Guidelines (WCAG) 2.0 are now also an ISO/IEC International Standard, ISO/IEC 40500:2012. In the context of WAI [32], an AT:

- relies on services (such as retrieving Web resources and parsing markup) provided by one or more other "host" user agents. AT communicate data and messages with host user agents by using and monitoring APIs.
- provides services beyond those offered by the host user agents to meet the requirements of users with disabilities. Additional services include alternative renderings (e.g. as synthesized speech or magnified content),

alternative input methods (e.g. voice), additional navigation or orientation mechanisms, and content transformations (e.g. to make tables more accessible).

The Digital Accessible Information System consortium [33] proposed standards (officially known as ANSI/NISO Z39.86 and ANSI/NISO Z39.98-2012) [34]. These standards define the format and the content of the electronic file set that comprises a digital talking book (DTB) and establishes a limited set of requirements for DTB playback devices. A DTB is a collection of navigable digital files arranged to present information to the target population via alternative media, namely, human or synthetic speech, refreshable braille, or visual display, e.g., large print.

In the case of scientific documents, accessibility deals also with mathematical or in general scientific symbols and formulas, as well as scientific graphics [35, 47].

MathML [36] is a W3C Recommendation defining an XML vocabulary for marking up mathematical expressions. It contains two sub-languages: presentation MathML (for describing the layout of a mathematical expression) and content MathML (for marking up some aspects of the meaning, or at least the mathematical structure, of expressions). Many software packages, including browsers, editors (e.g. MS-Word2010), and publishing software (including DAISY), have adopted MathML. MathPlayer [37] is freeware software that enhances Internet Explorer to display mathematical notation in large size or to transform any mathematical formula into spoken form using Text-to-Speech.

Scalar Vector Graphics [38] is a language for describing two-dimensional graphics and graphical applications in XML. SVGs offer a number of features to make graphics on the Web more accessible, to a wider group of users. The users who benefit include; those with low vision, color blind or blind users, and users of AT. A number of SVG features can also increase usability of content for many users without disabilities, such as users of mobile phones, tablets or other non-traditional Web access devices. SVG images are scalable, i.e. they can be zoomed and resized by the reader as needed. Scaling can help users with low vision and users of some AT (e.g., tactile graphic devices, which typically have low resolution).

Tactile graphics are simplified graphical images that use raised lines and textures to convey information. They can be produced either with tactile graphic software applications that send simple images to a Braille embosser or with dedicated devices on a microcapsule paper that is heated (“toasted”) to produce raised images.

4 Document-to-Audio

The primary goal of utilizing text signals or presentation elements in text documents is to distinguish parts of the text and to create a well-formed presentation of the content in order, for instance, to augment the reading performance or attract the reader. Authors use typography and layout in a specific way, e.g. there are “strict” typographic rules for the documents to be published in a scientific journal. But in newspapers and books the page designer (or the page manager), and not the author, has the primary responsibility for applying the typography and layout layers. Traditional factors that play a leading role in documents formation and presentation include readability (the gauge that measures how easily words, phrases, and blocks of text can be read) and legibility (the measure of easiness to distinguish one letter from another in a particular typeface). More recent factors include visual aesthetics and accessibility. Design attributes, such as character shapes and proportions, stroke weight and axis, affect legibility. Legibility is about perception. Readability refers to comprehension and visual comfort in reading long text passages [39]. People who design types, i.e. type designers, have a direct influence on legibility and people who set type, i.e. typographers, are determinant in text readability [40].

Thus, all the text signals, and particularly the typographic attributes, play an important role in accessing of a document by a print-disabled person or by those with an occasional or situational “disability”. Two typical application examples of the accessibility of text signals are: A) a blind individual usually access a book through the audio modality using a screen reader connected to a Text-to-Speech system. In order to achieve equal usability with the sighted reader of the same book, he requires from the TtS system to provide accessibility of text signals, i.e. font (type, size, color, etc.) and font style (such as bold, italics, underline) must be acoustically rendered. B) A car driver uses TtS to access his newspaper. A basic accessibility requirement for him is to know the significant of a title heading which is related with the font size used to represent the specific title in the text document.

Although the accessibility of the text signals is vital, almost all of the current Text-to-Speech systems do not take into account the semantics and the cognitive aspects of the presentation elements or text signals. Moreover, nowadays, there is not a systematic and formal methodology, based on sound research founding, on how to use typographic signals while creating an acoustic representation of a rich text document through TtS.

There is an effort towards Document-to-Audio (DtA) synthesis [4], which essentially constitutes the next generation of the Text-to-Speech systems. DtA supports the extraction of the semantics of document metadata of the presentation elements [41] and the efficient acoustic representation of both typography and text formatting, as well data tables [42]. This is accomplished in DtA through the modeling of the parameters of the synthesized speech signal by: (a) combining alternative text insertion in the document text stream, (b) altering the prosody, (c) switching between voices, and/or (d) inserting non-speech audio (like earcons) in the waveform stream, according to the class of metadata provided in the document.

We present here our Design-for-All based efforts towards rendering text signals to auditory modality using DtA synthesis at the typographic layer.

5. Acoustic Mapping of Visual Text Signals

First, we have applied analytics to large corpora of text documents in both English and Greek language (selected as a representative minor language with non-Latin alphabet) in order to extract knowledge for the logical, layout and typography layers embedded in text. Specifically, we statistically analyzed a corpus of 72 textbooks (a mixture of all subjects): 36 of them use by the K-12 schools in Greece and 36 in the English language used by the K-12 American Community School in Athens, Greece. The results indicate that in the case of text size has a range between 6 pts and 72 pts.

For the typography layer two approaches for rendering document signals to auditory modality are incorporated in DtA: a) direct mapping and b) emotional-based mapping.

5.1 Direct Mapping

Previous research in the field of direct acoustic mapping of font size signals in rich text documents involved the use of auditory icons [55, 56, 57], earcons [58, 59, 60], sound localization [61], interchange of speakers [57], as well as modification of pause duration and pitch range [62, 63].

In the present study our effort is to propose a mathematical relation between dimensions that are perceptible through different senses, such as the font-size and the loudness in TtS. Based on the relation similarity, each typographic cue is directly mapped into a respective acoustic cue. The principle of relational similarity explores two physical quantities with magnitudes that humans perceive by different senses in an analogous way. For example, the font size of a text and the loudness of the speech signal when the text is vocalized through TtS comprise a relational similarity in the case we perceive the change of their magnitudes in a proportional way.

51 Greek students, age 10-17 years, 29 females (15 blind and 14 sighted) and 22 males (10 blind and 12 sighted) were participating in the experiments.

The results show the following relation between the loudness (L) in db and font-size (F_s) in pts:

$$L = 0.6566 \cdot F_s + 45.7 \qquad (1)$$

This finding is consistent with the results of Smith and Sera [64].

5.2 Emotional-based Mapping

Present work introduces the emotional-based mapping methodology for rendering typographic signals to auditory modality. First, we measure and model the way emotional states are induced to the reader by the typographic signals. Then, following a number of psychoacoustic experiments, we determine analogous prosodic cues that produce the same emotional states to the listener when he/she hears the acoustic rendition of the document by a TtS system. Our ultimate goal is to incorporate automatic typography-to-speech mapping in DtA by emotional analogy between the visual (typographic signals) and the acoustic (speech prosody) modalities. In this framework we introduce the notion "typographic profile" of an emotion, in a similar way to the "prosodic profile". The way typographic attributes are used by a specific author or page designer constitute the typographic profile of the document. The profile defines the space within each emotion is located (and vice-versa). Figure 4 presents the prosodic and typographic profile respectively.

Recently [43], we have developed an automated reader's emotional state extraction process derived by the typographic cues, as well as an appropriate modeling of reader's emotional state response on document's typographic elements [44, 53] combined with the following expressive speech synthesis model [54]:

$$S = F \cdot \Delta E + I \qquad (2)$$

where:

$$S = \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_n \end{bmatrix} \quad F = \begin{bmatrix} a_1^P & a_1^A & a_1^D \\ a_2^P & a_2^A & a_2^D \\ \dots & \dots & \dots \\ a_n^P & a_n^A & a_n^D \end{bmatrix} \quad I = \begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{bmatrix}$$

$$\Delta E = \begin{bmatrix} P \\ A \\ D \end{bmatrix}$$

and:

S: the speech (prosodic) characteristics matrix
F: the factors matrix
I: the intercept (offset) matrix
ΔE: matrix of the emotional changes
P: Pleasure in [−100, 100]
A: Arousal in [−100, 100]
D: Dominance in [−100, 100]

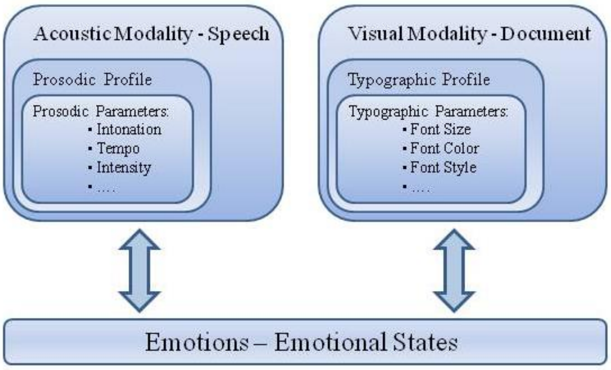


Figure 4: The prosodic and typographic profiles

In the PAD space (Pleasure, Arousal and Dominance), we can model the emotional changes induced to the reader by text size using the following equation:

$$\Delta E_{\{d\}}(s) = B_3^{\{d\}} \cdot s^3 + B_2^{\{2\}} \cdot s^2 + B_1^{\{d\}} \cdot s + I_{nt}^{\{d\}} \quad (3)$$

where:

d: P, A or D
P: Pleasure in [-1, 1]
A: Arousal in [-1, 1]
D: Dominance in [-1, 1]
B_1, B_2 and B_3 the polynomial coefficients
I_{nt}: Intercept
s: font size in px.

Table 1 presents the polynomial coefficients for the above equation (3) that relates font size with the changes in the PAD emotional states. Moreover, the standard errors of these polynomial coefficients are presented in parenthesis in Table 1.

Table 1: Polynomial coefficients used in equation (3) for modeling the PAD emotional changes induced by the with text size. Standard errors are given in parenthesis

	Pleasure	**Arousal**	**Dominance**
$\mathbf{B_1}$	0.25 (0.033)	-0.59 (0.12)	-0.41 (0.13)
$\mathbf{B_2}$	-0.0052 (0.0008)	0.029 (0.006)	0.016 (0.007)
$\mathbf{B_3}$	0 (0)	-0.00043 (0.0001)	-0.0002 (0.0001)
$\mathbf{I_{nt}}$	-2.45 (0.31)	3.58 (0.74)	2.85 (0.77)

Applying the data of Table 1 in equation (3) we model the PAD emotions induced to the readers by font size using the following equations:

$$P = -0.0052 \cdot s^2 + 0.25 \cdot s - 2.45$$

$$A = -0.00043 \cdot s^3 + 0.029 \cdot s^2 - 0.59 \cdot s + 3.58$$

$$D = 0.0002 \cdot s^3 + 0.016 \cdot s^2 - 0.41 \cdot s + 2.85$$

Thus, the matrix of the emotional changes *ΔE* and the intercept matrix *I* in equation (2) for the case of signals derived from the font size are:

$$\Delta E = \begin{bmatrix} 0.25 & -0.0052 & 0 \\ -0.59 & 0.029 & -0.00043 \\ -0.4 & 0.016 & 0.0002 \end{bmatrix}$$

$$I = \begin{bmatrix} -2.45 \\ 3.58 \\ 2.85 \end{bmatrix}$$

The above results are consistent with those of Bernardini et al. [65], who noted that the preference emotional level decreases not only using smaller font size but also using larger font size. Moreover, Bernard et al. [66] showed that text at 12 pt size on a computer display is significantly preferred compared to the 10 pt size.

6 Conclusions

We have presented our recent efforts towards advanced document accessibility by supporting not only the accessibility of the text content of a document, but also the accessibility of the text signals, i.e. the writing devices that emphasizes aspects of a text's content or structure. First, we have introduced text signals, along with an appropriate document's architecture, and the main principles and technological topics of document accessibility. Then, the emerging technological approach of Document-to-Audio (DtA) we have developed was presented. DtA essentially constitutes the next generation of Text-to-Speech systems and is capable to provide much better accessibility to documents. Finally, we introduced the emotional-based mapping methodology for rendering typographic signals to the auditory modality for the case of font size.

Our future research include a number of experiments with blind and sighted participants in order to study the appropriate relation similarities of direct mapping, as well as the emotional based mapping, of other typographic signals (e.g. font type or color and font style, such as bold, italics, underline) to auditory modality in DtA. Moreover, we plan a psychoacoustic experiment to compare the direct mapping approach and the emotional based mapping in DtA.

Acknowledgements

This work has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) under the Research Funding Project: “THALIS-University of Macedonia- KAIKOS: Audio and Tactile Access to Knowledge for Individuals with Visual Impairments”, MIS 380442.

References:

[1] D.S Doermann, E. Rivlin and A. Rosenfeld, The function of documents. *Image Vision Computing*, Vol. 16, No. 11, 1998, pp. 799–814.

[2] K. Fellbaum and G. Kouroupetroglou, Principles of Electronic Speech Processing with Applications for People with Disabilities. *Technology and Disability*, Vol. 20, No. 2, 2008, pp. 55–85.

[3] M. McLuhan, and Q. Fiore, *The Medium is the Message*. Berkeley, CA: Gingko Press, 2005.

[4] G. Kouroupetroglou and D. Tsonos, Multimodal Accessibility of Documents. In S. Pinder (Ed.), *Advances in Human-Computer Interaction*. Vienna: I-Tech Education and Publishing, 2008, pp. 451–470.

[5] R.F. Lorch, Text-Signaling Devices and Their Effects on Reading and Memory Processes. *Educational Psychology Review*, Vol. 1, No. 3, 1989, 209–234.

[6] J. Lemarié, H. Eyrolle, and J.M. Cellier, Visual signals in text comprehension: How to restore them when oralizing a text via a speech synthesis? *Computers in Human Behavior*, Vol. 22, No. 6, 2006, pp. 1096–1115.

[7] R.F. Lorch, H.T. Chen, and J. Lemarié, Communicating Headings and Preview Sentences in Text and Speech. *Journal of Experimental Psychology Applied*, Vol. 18, No. 3, 2012, pp. 265–276.

[8] CRPD, Convention on the Rights of Persons with Disabilities of the United Nations. http://www.un.org/disabilities/

[9] ADA, Americans with Disabilities Act., http://www.ada.gov/

[10] Section 508. Section 508 of the Rehabilitation Act. http://www.section508.gov

[11] i2010, A European Information Society for growth and employment. http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm

[12] Europe 2020, Europe 2020: the European Union's ten-year growth strategy. http://ec.europa.eu/europe2020/

[13] C. Stephanidis, G. Salvendy, D. Akoumianakis, N. Bevan, J. Brewer, P.L. Emiliani, et al., Toward an information society for all: An international R&D agenda. *International Journal of Human-Computer Interaction*, Vol. 10, 1998, pp. 107-134.

[14] WebAIM, Rich Media Accessibility. http://webaim.org/techniques/

[15] Microsoft, *Accessibility in Microsoft Office 2010*. http://www.microsoft.com/enable/products/office2010/

[16] Adobe, *Acrobat accessibility training resources*. http://www.adobe.com/accessibility/products/acrobat/training.html

[17] ZHAW, Plug-ins for accessible PDF documents. http://www.init.zhaw.ch/en/engineering/institute-of-applied-information-technology/focus-areas/human-information-

interaction/plug-ins-for-accessible-pdf-documents.html
[18] A. Cook, and S. Hussey, *Assistive Technologies: Principles and Practice*. Mosby, 2001.
[19] D. Freitas and G. Kouroupetroglou, Speech Technologies for Blind and Low Vision Persons. *Technology and Disability*, Vol. 20, No. 2, 2008, 135–156.
[20] M. Scherer and R. Glueckauf, Assessing the benefits of Assistive Technologies for activities and participation. Rehabilitation Phychology, Vol. 50, No. 2, 2005, pp. 132–141.
[21] World Health Organization, *International classification of functioning, disability and health (ICF)*. Geneva, Switzerland, 2001.
[22] P. Blenkhorn, G. Evans, A. King, S. Hastuti Kurniawan, and A. Sutcliffe, Screen magnifiers: evolution and evaluation. *IEEE Computer Graphics and Applications*, Vol. 23, No. 5, 2003, pp.54–61.
[23] G.L. Goodrich and R.D. Quillman, CCTVs: choices and considerations. *Journal of Visual impairment and Blindness*, Vol. 73, No. 2, 1978, pp. 386-398.
[24] G.E. Fonda, H. Thomas, and R.N. Schnur, Evaluation of closed-circuit television as an optical aid for the low vision patient. *Trans Am Acad Ophthalmol Otolaryngol* Vol. 79, 1975, pp. 468–482.
[25] G. Kouroupetroglou, *Assistive Technologies and Computer Access for Motor Disabilities*. Pennsylvania, Hershey PA, USA. Information Science Reference Press, IGI Global, 2013.
[26] G. Kouroupetroglou, *Disability Informatics and Web Accessibility for Motor Limitations*. Pennsylvania, Hershey PA, USA. Information Science Reference Press, IGI Global, 2013.
[27] G. Weber, Reading and pointing-new interaction methods for Braille displays", chapter in Edwards, A. *Extra-ordinary human-computer interaction*, New York, USA, Cambridge University Press, 1995.
[28] V. Luceno, Software and Hardware: DAISY Players. http://www.daisy.org/ software-players
[29] A. Pino, G. Kouroupetroglou, H. Kacorri, A. Sarantidou, and D. Spiliotopoulos, An Open-Source / Freeware Assistive Technology Software Inventory. *Lecture Notes in Computer Science*, Vol. 6179, 2010, pp. 178-185.
[30] ATHENA, Open-Source/Freeware Assistive Technology Software Inventory. http://access.uoa.gr/ATHENA
[31] WAI, Web Accessibility Initiative. http://www.w3.org/WAI/
[32] W3C The World Wide Web Consortium. http://www.w3.org/
[33] DAISY, The Digital Accessible Information System Consortium. http://www.daisy.org/
[34] ANSI/NISO, ANSI/NISO Z39.98-2012. http://www.daisy.org/z3998/2012/z3998-2012. html
[35] G. Kouroupetroglou, and H. Kacorri, Deriving Accessible Science Books for the Blind Students of Physics. In A. Angelopoulos (Ed.), *7th International Conference of the Balkan Physical Union*, Vol. 1203, No. 1, 2010, pp. 1308-1313. Melville, NY: American Institute of Physics.
[36] MathML, Mathematical Markup Language. http://www.w3.org/Math/
[37] MathPlayer, http://www.dessci.com/en/ products/mathplayer
[38] SVG, Scalable Vector Graphics. http://www.w3.org/Graphics/SVG/
[39] Y.C. Tai J. Sheedy, and J. Hayes, Effect of letter spacing on legibility, eye movements, and reading speed. *Journal of Vision*, Vol. 6, No. 6, 2006, article 994. doi: 10.1167/6.6.994.
[40] C.Y. Suen, N. Dumont, M. Dyson, Y.C. Tai, and X. Lu, Evaluation of Fonts for Digital Publishing and Display. Proceedings of the *International Conference on Document Analysis and Recognition*, 2011, pp. 1424–1436.
[41] F. Fourli-Kartsouni, K. Slavakis, G. Kouroupetroglou, and S. Theodoridis, Bayesian Network Approach to Semantic Labeling of Text Formatting in XML Corpora of Documents. *Lecture Notes in Computer Science*, Vol. 4556, 2007, pp. 299–308.
[42] D. Spiliotopoulos, G. Xydas, G. Kouroupetroglou, V. Argyropoulos, and K. Ikospentaki, Auditory accessibility of data tables using naturally derived prosody specification. *Universal Access in the Information Society*, Vol. 9, No. 2, 2010, pp. 169–83.
[43] G. Kouroupetroglou, D. Tsonos, and E. Vlahos, DocEmoX: A System for the Typography-Derived Emotional Annotation of Documents. *Lecture Notes in Computer Science*, Vol. 5616, 2009, pp. 550–558.
[44] D. Tsonos, and G. Kouroupetroglou, Modeling readers' emotional state response on documents' typographic elements. *Advances in Human Computer Interaction*, 2011, pp. 1–18. doi:10.1155/2011/206983
[45] C. Stephanidis, Editorial. *Universal Access in the Information Society*, Vol. 1, 2001, pp. 1–3.

[46] M. Malcangi and P. Grew, Toward language-independent text-to-speech synthesis, *WSEAS Transactions on Information Science and Applications*, Vol. 7, No. 3, 2010, pp. 411–421.

[47] N. Isaila, and I. Smeureannu, The access of persons with visual disabilities at the scientific content, *WSEAS Transactions on Computers*, Vol. 9, No. 7, 2010, pp. 788–797.

[48] O. Pavel, Automatic System for Making Web Content Accessible for Visually Impaired Users. *WSEAS Transactions on Computers* Research, Vol. 1, No. 2, 2006, pp. 325–328.

[49] J.H. Spyridakis, Signaling effects: A review of the research—Part I. *Journal of Technical Writing and Communication*, Vol. 19, 1989, pp. 227–240.

[50] J. Lemarié, R.F. Lorch, H. Eyrolle and J. Virbel, SARA: A Text-Based and Reader-Based Theory of Text Signaling. *Educational Psychologist*, Vol. 43, 2008, pp. 27–48.

[51] R.F. Lorch, J. Lemarié and R.A. Grant, Signaling hierarchical and sequential organization in expository text. *Scientific Studies of Reading*, Vol. 15, 2011, pp. 267–284.

[52] Z.H. Han, E.S. Park and C. Combs, Textual Enhancement of Input: Issues and Possibilities. *Applied Linguistics*, Vol. 29, 2008, pp. 597–618.

[53] D. Tsonos, G. Kouroupetroglou and D. Deligiorgi, Regression Modeling of Reader's Emotions Induced by Font Based Text Signals. *Lecture Notes in Computer Science*, Vol. 8010, 2013, pp. 434-443.

[54] M. Schröder, Expressing degree of activation in synthetic speech. *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14, No. 4, 2006, pp. 1128–1136.

[55] W. Gaver, Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, Vol. 2, 1986, pp. 167- 177.

[56] E.D. Mynatt and W.K. Edwards, Mapping GUIs to auditory interfaces. *Proceedings of the ACM: UIST '92*, New York, 1992, pp. 61-70.

[57] F. James, Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext. *Proceedings of the International Conference on Auditory Display (ICAD)*, 1997, pp. 97-103.

[58] M. Blattner, D. Sumikawa and R. Greenberg, Earcons and icons: Their structure and common design principles. In Glinert, E.P. (ed.). *Visual Programming Environments: Applications and Issues*. IEEE Computer Society Press, Los Alamitos, CA, 1990, pp. 582-606.

[59] S. Brewster, P. Wright and A.D.N. Edwards, An evaluation of earcons for use in auditory human-computer interfaces. *Proceedings of INTERCHI '93*, Amsterdam, 1993, pp. 222-227.

[60] T.V. Raman, *Audio system for technical readings*. Sringer-Verlag, 1994.

[61] K. Kallinen, *Using sounds to present and manage information in computers*. Center for Knowledge and Innovation Research, Helsinki School of Economics, Finland, 2003.

[62] H.J.N. den Ouden, L.G.M. Noordman and J.M.B. Terken, The prosodic realization of organizational features of texts. *Proceedings of the speech prosody conference*, 2002, pp. 543–546.

[63] S. Brewster, *Providing a model for the use of sound in user interfaces*. PhD Theses, Department of Computer Science University of York, Heslington, 1991.

[64] L. Smith, and M. Sera, A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, Vol. 24, 1992, pp. 99–142.

[65] C. Bernardini, V. Ambrogi, G. Fardella, L. Perioli, and G. Grandolini, How to improve the readability of the patient package leaflet: a survey on the use of colour, print size and layout. *Pharmacological Research*, Vol. 43, No. 5, 2001, pp. 437–443.

[66] M. L. Bernard, B. S. Chaparro, M. M. Mills, and C. G. Halcomb, Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *International Journal of Human Computer Studies*, Vol. 59, No. 6, 2003, pp. 823–835.