Modeling Acoustic Rendition of Documents' Typography using Expressive Speech Synthesis for Sighted and Blind Users

Dimitrios Tsonos*

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
dtsonos@di.uoa.gr

Abstract. The accessibility to printed and electronic documents (books, newspapers, web content) by the print disabled, as well as the moving users and the elderly, is based on the possibility to convert them (in real time) into acoustic and/or haptic modality. Besides its content, a printed or electronic text document contains a number of presentation visual elements that apply design glyphs or typographic elements, such as font (type, size and color) and font style (bold and italics). Regardless the important progress achieved in Text-to-Speech systems, they do not support the efficient sonification of the semantics and cognitive aspects of the Visual Presentation Elements in Documents (VPED). Essentially all this additional metadata vanishes during document's processing towards its acoustic or haptic rendition.

This dissertation deals with the sonification of the VPED metadata during their transformation to speech. The approach includes: a) the automatic extraction of VPED induced emotional states to the reader and b) their acoustic rendition using expressive emotional synthetic speech.

Focusing on the development of a system for the automatic extraction of the VPED induced emotional states and the appropriate document annotation, a novel architecture is proposed for the multimodal universal accessibility of documents, regardless of their natural language, content and culture. A quantitative model is developed for the sonification of the typographic alterations by: i) the mathematical formulation of the induced reader's emotional state, based on the dimensional nature of the emotions ("Pleasure", "Arousal" and "Dominance"), and ii) their mapping into prosodic alterations of the expressive synthetic speech.

The prosodic model was evaluated, using psychoacoustic experiments, whether the listeners can acoustically recognize the typographic alterations. The results were positive even in the case of listeners without any previous training. Furthermore, the evaluation of the developed model by sighted and blind students of primary education shows enhancement of their performance during the didactic process.

Keywords: Human Computer-Interaction, Universal Accessibility, Design-for-All, Emotions, Expressive Speech Synthesis

* Dissertation Advisor: Georgios Kouroupetroglou, Assoc. Professor

1 Introduction

A document is the "medium" in which a "message" (information) is communicated [1]. The term "signal" is introduced as, "the writing device that emphasize aspects of a text's content or structure without adding to the content of the text" [2]. It attempts to pre-announce or emphasize content and/or reveal content relationship [3] [4]. The title, heading, typographic cues are considered as signals. 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) [5]. Tsonos and Kouroupetroglou [6] categorize the signals, focusing on visual presentation of documents, into three layers: logical, layout and typographic. All the devices [2], either mentioned as signals or layers:

1. share the goal for directing the reader's attention during reading.
2. facilitate specific cognitive process occurring during reading.
3. ultimate comprehension of text information.
4. may influence memory on text.
5. direct selective access between and within texts.

People with print disabilities (as well as moving and/or elderly) require printed or electronic documents in alternative formats, such as Braille, audio, large print or electronic text. Text-to-Speech (TtS) is a common software technology that converts in real-time any electronic text into speech [7]. Most of the current Text-to-Speech systems do not include effective provision of the semantics and the cognitive aspects of the visual (e.g. font style and size) and non-visual signals (e.g. emphasis).

The acoustic rendition of documents' visual information is a complex procedure. In order to accomplish this task, it is used methodologies such as specific sounds (earcons) before and/or after the text that is presented in a specific typographic manner or the use of the prosodic variations in speech synthesis. This method can be considered as "direct acoustic rendition" of the typographic elements because of the direct link/mapping between typography and prosody. Recently, there was an effort towards Document-to-Audio (DtA) synthesis [34] [35]. It essentially constitutes the next generation of the Text-to-Speech systems, supporting the extraction of the semantics of document metadata [8] and the efficient acoustic representation of text formatting [9-12] through modeling the parameters of the synthesized speech signal [13] [14], considered as "indirect acoustic rendition".

W3C introduces Aural Cascaded Style Sheets [15] recommendations towards the rendition of web documents typographic elements using synthetic speech. Kallinen et al. [16] introduces the term "auditive boldfacing". They examine the effects of auditive "boldfacing", i.e., occasional lowering of the voice by two semitones, on the memory performance of audio business news differing in terms of valence and arousal in 28 subjects. "Audio boldfacing" enhanced the immediate memory performance but decreased the longer term performance. Truillet et al. [17] present an experimental study towards evaluating "sound fonts". They study the acoustic rendition of salient words using speech synthesis, obtained either presented by addition of a verbal description of the typographic attribute ("in bold" with a decrease of 15% of the current pitch is said before the salient word) or pronounced with an increase of 13% of the default pitch. Through various psycho-acoustic manipulations (pitch, volume and speed variations of synthetic speech), Argyropoulos et al. [18] examine their effectiveness for the understanding of specific information (typographic attributes - bold and italic) by thirty sighted and thirty blind participants. A preliminary study of auditory rendition of typographical and/or punctuation information using expressive speech synthesis is presented in [14]. The aim is to increase the expressiveness of the already existing TTS system of France Telecom using prosodic rules. Four prosodic elements are proposed for use: pitch, rate, volume and break. The acoustic rendition of LATEX documents was proposed by Raman [19].

The above studies simply propose rules for the implementation of the acoustic rendition of specific typographic elements. There is not a systematic approach towards the acoustic rendition of typographic elements. So, it is difficult to implement an automated system, which supports this functionality. The present dissertation aims, not only to propose a set of rules, but also a methodology that through its implementation we are able to create a system that it can automatically produce and apply the typography to acoustic mapping rules utilizing emotions and/or emotional states.

Thus, this work:

1. presents a novel integrated XML-based system, content, language and domain independent, for the real-time production, presentation and navigation of multimodal accessible documents by conforming to international guidelines and standards.
2. proposes a modular, content-free and language independent methodology, for the acoustic rendition of document's typographic elements, using the dimensional approach of emotions (this assumption is based on the universal character of emotions across languages and cultures).
3. introduces a set of rules towards the automated reader's emotional state response on document's typographic elements, based on the dimensional description of emotions.
4. introduces specific rules for the acoustic rendition of font style and size alterations, combining reader's emotional state variations and expressive speech synthesis.
5. evaluates the derived acoustic mapping rules, through listeners' preferences and the level of recognition (psychoacoustic experiments).

2 Multimodal Accessibility of Documents

The discussion on accessibility of documents imposes questions on how the documents can be accessed either in visual, acoustic or haptic modality. Some studies are trying to create accessible documents in the acoustic modality using speech synthesis or by combining earcons, auditory icons and 3D sounds for the auditory [20]. In this section a novel integrated XML-based system is presented, for the real-time production, presentation and navigation of multimodal accessible documents by conforming to international guidelines and standards (such as W3C's and ANSI/NISO). This approach includes a unified methodology for the multimodal rendering of text formatting, text structure, text layout and non-textual elements.

The overall proposed architecture consists of two main parts (Fig. 1): the User Interface and the Document Explorer. The later is responsible for document analysis during the *production process* and the exploration in the documents (during *navigation task*). User Interface is responsible for the multimodal interaction with the documents. It collects the user preferences and the device requests, as well as the navigation commands, and executes the *presentation task*. These parts are implemented by two different modules following the Client – Server model. The Document Explorer can be hosted on a powerful server machine due to the resource demanding tasks that performs. In contradiction, the User Interface can be hosted on any common computer (e.g. a personal computer, a PDA, mobile smart phone). This kind of implementation fulfills the Web Content Accessibility Guidelines, in order to be device and software independent. The implementation and the communication between the modules are XML-based. The proposed architecture also can be slightly altered and can be used e.g. in classroom for sighted and blind students [27] [28] [29].

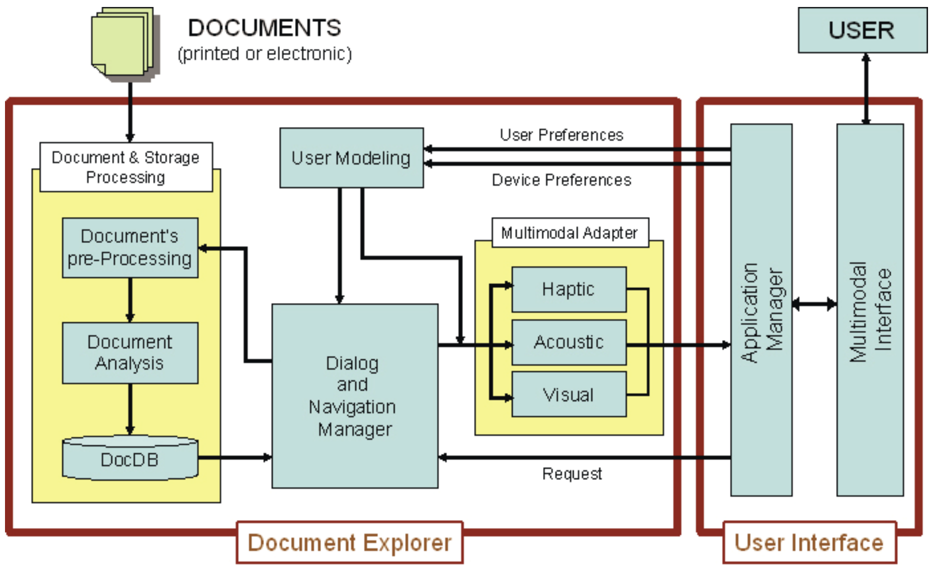

Fig. 1. The XML-based system for multimodal accessibility of documents

3 Acoustic Rendition of Documents' Typography using Expressive Speech Synthesis

While reading a document, time passes and typographic elements alter, thus reader's emotions and emotional state vary. It is a similar syllogism to one used in the expressive speech synthesis model by Schröder [23] who uses the dimensional approach of emotions. It is worthy to mention that the ordinary alterations of typographic elements are not extreme e.g. font size from 12pt to 14pt. or plain text to italized. Thus the dimensional approach of emotions, either in typography or speech, is considered as an appropriate approach for the mapping of the typographic elements to prosodic variations. The dimensional approach of emotions in speech is based on the continuous nature of the emotion representation and variation, so we are able to utilize this model in order to represent the analogous behavior of typographic elements in documents to the acoustic modality.

3.1 Visual and Acoustic Modality Analogy

The primary goal of using presentation elements and specifically typographic elements in text documents, is to distinguish parts of the text and to create a well formed presentation of the content in order e.g. to augment the reading performance, attract the reader, render semantics through the visual channel. Authors use typography in a specific way, e.g. in scientific documents there are “strict” typographic rules. But in newspapers and magazines, the page-art design department and not the authors has the primary responsibility on applying typography. The notion “typographic profile” of an emotion is introduced, in a similar way that “prosodic profile” is presented, in [24]. The way typographic parameters are used (here are referred as “typographic elements”), constitute the typographic profile of the document. The profile defines the space within each emotion is located (and vice-versa). In Figure 2 are presented the resemblances of prosodic and typographic profile respectively.

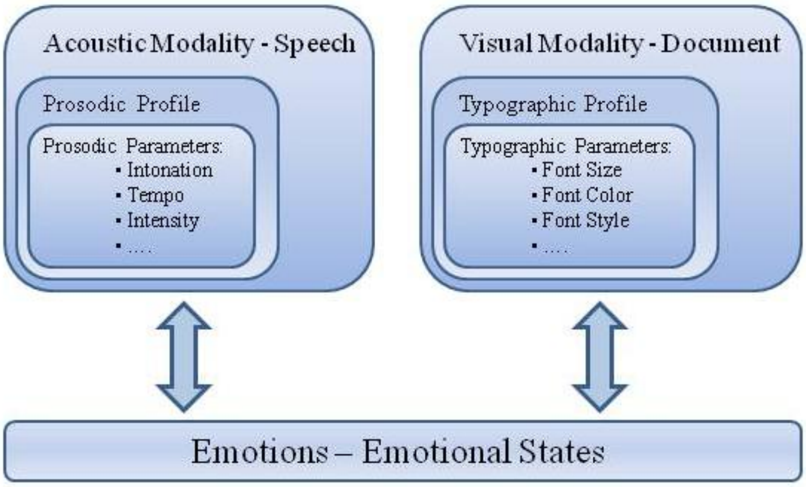

Fig. 2. The prosodic and typographic profiles.

Typography-to-emotions and emotions-to-speech approaches, can map the typographic alterations into speech prosodic variations and we are able to observe which speech parameters are influenced and how they vary, while changing the font style/size attributes. The combination of these models provides a model on how the prosody varies according to typographic alterations. Thus, human emotions constitute the medium for the two channels / modalities to interact - communicate.

3.2 Modeling Reader's Emotional State Response

In [6] are described the results of an experimental study towards modeling the reader's emotional state variations derived by the typographic elements in documents. Based on the experimental results, it is proposed a set of mapping rules on how typographic elements, such as typesetting (plain, italics, bold and bold-italics), and font elements (font/background color combinations, font type and font size) affect the reader's emotional state. The study was based on the dimensional approach of the emotions (namely "Pleasure", "Arousal" and "Dominance") and the experimental procedure was designed and implemented according to IAPS experimental guidelines, using the Self Assessment Manikin test [21]. It is clear that emotional states have similar affection and variations across language and culture. This assumption derives from the comparison of our experimental results with those from similar studies for other languages and cultures.

In the present dissertation is introduced a model for the quantitative representation of the results. It is used the polynomial regression for the case of font size variations (continuous values) and simple percentage variations for the case of font type and typesetting (discrete values). Based on this modeling and the need of implementing the rules for the automated reader's emotional state extraction [22] [33], a generic mathematical description is proposed on how the variations of the typographic elements can be mapped into emotional state values.

Some major findings of this work are: Three polynomial equations describe the way the emotional dimensions are affected by the font size (Table 1). While font size increases up to 26px, "Pleasure" dimension increases respectively. From this value and higher there is a decrease. In contradiction, "Arousal" and "Dominance" decreases up to 15px and 18px respectively. Beyond these values there is an increase in both dimensions. The results denoted that only "Pleasure" is affected by the font type. "Times New Roman" are considered more pleasant than Arial (23% increase of "Pleasure") and only "Pleasure" is affected by typesetting elements.

Table 1. Constants of the modeling of PAD relationship on text size using equation (1) with their standard errors.

	Pleasure		Arousal		Dominance	
		SE		SE		SE
Intercept	-2.45287	0.30771	3.58459	0.73588	2.85307	0.77138
B_1	0.24942	0.03299	-0.59264	0.12133	-0.41265	0.12828
B_2	-0.00523	0.00077	0.02866	0.00632	0.01631	0.0067
B_3	0	0	-0.000427111	0.0001029	-0.000199863	0.0001093
Adjusted R-Square	**0.90687**		**0.78103**		**0.84705**	

3.3 Typography to Expressive Speech Synthesis Modeling

Following the proposed methodology and utilizing the model proposed by Tsonos & Kouroupetroglou [6], we can calculate the emotional state variations from the neutral state derived by the "typographic baseline". The typographic baseline is defined as "the most frequent value of the typographic element that is occurred in the whole document". A statistical survey [25] [32] was conducted on printed school book and newspapers. The more frequent font sizes that consist the baseline of the documents are 10pt, 12pt and 14pt (throughout all books and newspapers). Also it is observed that the increment of font size (occurred e.g. in title, headline) has values of 12pt, 14pt, 16pt, 18pt, 20pt, 22pt and decrement (e.g. footnote) at 8pt, 9pt, 10pt, 11pt, 12pt, 13pt depending the font size baseline.

The mapping of typographic alterations into prosodic variations can be accomplished combining the results derived from the reader's emotional state response model and the expressive speech synthesis model proposed by Schröder [23] and its implementation, the XSLT file used in OpenMary System [26]. The prosodic variations are very small due to little variations of the emotional states derived from the alterations of the font size from the baseline. These variations are not distinct to the human listener. So it is needed to be "augmented" in order to be more comprehensible. Thus, it is proposed the normalization of the values using linear quantization of these variations.

The levels of quantization are defined calculating the difference between minimum and maximum values for each prosodic element and dividing it by 6 levels of increment or decrement. The number of levels is considered as the optimum and derives from observations while trying not to have duplicate cases for each prosodic element variations. Greater and lower values, is observed many duplicate cases.

The value associated for each level of increment/decrement for each prosodic element is:

- 12% for pitch (corresponds approximately to 2 semitones of increment/decrement).
- 10% - 14wpm for speech rate (6 levels of increment correspond to maximum rate of 224 wpm not exciding the upper limit of 225 wpm).
- 15% for intensity corresponding to more than >1dB increment/decrement.
- After and before each prosodic variation there is a speech pause of 150 ms duration.

The model proposes multiple cases for each typographic alteration (3 cases for each font size and 2 cases for each font style alteration from the baseline). Using many prosodic variations in speech and implementing the same typographic element in different ways, the listener can be confused and annoyed. Also, in the present study, we want to create a model that can be applied to any document, examining only the typographic alterations from the baseline and not absolute values. Thus, a psychoacoustic experimental procedure was implemented for the selection of the optimum acoustic rendition of typographic alterations. The results are [30]:

1. plain to bold alteration affects pitch (-12%) and rate (-10%),

2. plain to italics, affects pitch (+36%) and rate (+30%),
3. plain to bold-italics, affects pitch (-36%) and rate (-30%),
4. increasing font size (+2pt) should increase pitch (+12%), decrease rate (-10%) and volume (-15%),
5. decreasing font size (-1pt) should increase pitch (+12%), rate (+10%) and volume (+30%).

4 Prosodic Model Evaluation

In order to evaluate the prosodic model, another experimental procedure was conducted [30]. It is examined whether the listeners are able to recognize the typographic alterations (from plain to bold, italics, bold-italics, increment and decrement of font size) by their corresponding prosodic variations.

Eleven male and eight female participated in the experimental procedure (total 19 participants) - aging from 20 to 32 year-old (Mean Age = 27.8, SD = 3.2). They were undergraduate or postgraduate students from our department; their native language is Greek, naive to any previous experiment. Totally, 20 stimuli were used in Greek language, 12 for the font style elements and 8 for the font size. Each stimulus contained a sentence that is acoustically expressed in a typographic element or a phrase of the sentence consisting from 1, 2 or 3 words. The sentences were converted into speech using the Document-to-Audio platform [12].

There was *no training session* for the participants to be familiarized with prosodic variations due to typographic (font style and size) alterations. They participated each one in two sessions: evaluation of the acoustic rendition of a) font style elements and b) font size elements respectively. The stimuli were displayed simultaneously with the sentence without the typographic elements. The evaluators were asked which typographic element they believe is associated with the sentence or phrase (font style: bold, italics, bold-italics or none / font size: size increment, size decrement or none)

The results denote that participants easily recognized the font style and size variations. In details:

- Italics were recognized successfully.
- Bold-italics failed to be recognized by the evaluators and were confused with the bold (almost all cases were assessed by the majority of the participants as bold). This is due to the similar prosodic variations for the both cases of bold and bold-italics (decrement of pitch and rate).
- Font size variations are also recognized successfully. In one case the correct answers were marginally lower 36.8%. We estimate that probably the false answer is due to stimulus implementation (e.g. the quality of speech synthesis for the specific sentence).

The model was also evaluated through another experimental procedure [31] by 10 sighted and 10 blind students of primary education (Mean Age=11 years and 6 months). The visual stimuli (MS PowerPoint slides) were presented using an interactive whiteboard. The content of the slides were converted into speech using

DEMOSTHeNES TtS applying the proposed prosodic model. The results show enhancement of their performance during the didactic process.

5 Conclusions

To ensure that people with print disability are able equally to participate in society, it is crucial to develop more effective ways for the accessibility of both printed and electronic documents. In this study it is presented an integrated architecture on how a document is structured. Based on international standards and guidelines, it is proposed a novel holistic XML-based system for the real time production, presentation and navigation of multimodal accessible documents.

Also it is presented a novel approach for the acoustic rendition of typographic alterations in documents targeting to a methodology that is modular, content-free and language independent. Based on reader's emotional state variations on documents typographic alterations (from plain text to bold, italics and bold-italics / several font sizes) and using expressive speech synthesis we are able to acoustically render these alterations in order to be clearly distinguished by the human listener. This description can be further extended to other modalities, to visual and/or haptic, with prior knowledge of their mathematical modeling.

Utilizing reader's emotional state model proposed by Tsonos and Kouroupetroglou [6], the difference of the percentage emotional state variations (Pleasure, Arousal and Dominance) is calculated for a set of typographic alterations (from plain text to bold, italics and bold-italics / font size alterations from specific font size baselines, that was studied in the statistical survey presented in [25] [32]). Utilizing expressive speech synthesis [23] we are able to map these emotional state variations into prosodic (pitch, rate and volume) ones. Using this "indirect" manner, we map the typographic alterations into prosodic variations. The resulting variations are below the level of human listener's perception and performance. Thus, in order to enhance the perception, the normalization is proposed through linear quantization of the prosodic variations and assign to each level a minimum perceived value for each prosodic element. Because the resulting acoustic model sets multiple ways for the acoustic rendition of font style and size alterations, an experimental procedure was conducted in order to select the optimum implementation.

In general, it is observed that the recognition results are promising. Some major findings are: a) plain to bold alteration affects pitch (-12%) and rate (-10%), b) plain to italics affects pitch and rate (increasing them by 36% and 30% respectively), c) increasing font size (+2pt) should increase pitch (+12%), decrease rate (-10%) and volume (-15%), d) decreasing font size (-1pt) should increase pitch (+12%), rate (+10%) and volume (+30%). The results denoted the lack of representing bold-italics (but it has a similar behavior with the case of bold, decrement of both pitch and rate). Also we should have in mind that the participants were *not trained* to the acoustic variations and the acoustic representation of typographic elements. So it would be very interesting to examine, in a future study, the proposed methodology in an experimental procedure with trained participants to these prosodic variations. Another proposed future study is to change the minimum perceived levels for pitch, rate and volume in order to eliminate the pitfalls of the current study (e.g. the misunderstanding of bold-italics and bold during their acoustic rendition) or the use of non-linear quantization instead of the proposed linear.

Also it would be very interesting the study of acoustic rendition of typographic elements in a similar methodology using discrete emotions (than the dimensional approach). This would extend the use of current methodology in systems than can process reader's discrete emotions derived by the typographic elements/alterations and the T-t-S systems that only support discrete emotions approach. This study is not limited to the acoustic rendition of typographic elements but it can be extended for the acoustic rendition of emotions that derive from the content itself. Finally, the proposed methodology is targeting to other modalities (multimodal approach). It sets the guidelines towards future work for the haptic modality or for a multimodal system due to its modular architecture.

6 References

1. McLuhan, M., Fiore, Q.: The Medium is the Message. Berkeley, USA, Gingko Press (2005)
2. Lorch, R.F.: Text-Signaling Devices and Their Effects on Reading and Memory Processes. Educational Psychology Review, 1, 209-234 (1989)
3. Spyridakis, J.H.: Signaling effects: A review of the research—Part I. Journal of Technical Writing and Communication. 19, 227-240 (1989)
4. Lemarie, J., Eyrolle, H., Cellier, J. M.: Visual signals in text comprehension: How to restore them when oralizing a text via a speech synthesis?. Computers in Human Behavior, 22, 1096-1115 (2006)
5. Han, Z.H., Park, E.S., Combs, C.: Textual Enhancement of Input: Issues and Possibilities. Applied Linguistics, 29, 597-618 (2008)
6. Tsonos, D., Kouroupetroglou, G.: Modeling reader's emotional state response on document's typographic elements. Advances in Human Computer Interaction, 2011, 1-18, (2011)
7. Fellbaum, K., Kouroupetroglou, G.: Principles of Electronic Speech Processing with Applications for People with Disabilities, Journal Technology and Disability, 20(2), 55-85 (2008)
8. Fourli-Kartsouni, F., Slavakis, K., Kouroupetroglou, G., Theodoridis, S.: A Bayesian Network Approach to Semantic Labelling of Text Formatting in XML Corpora of Documents. LNCS, vol. 4556, pp. 299-308 (2007)
9. Xydas, G., Kouroupetroglou, G.: Augmented Auditory Representation of e-Texts for Text-to-Speech Systems, LNAI, vol. 2166, pp. 134-141 (2001)
10. Xydas, G., Kouroupetroglou, G.: Text-to-Speech Scripting Interface for Appropriate Vocalisation of e-Texts. In: EUROSPEECH 2001, pp. 2247–2250 (2001)
11. Xydas, G., Spiliotopoulos, D., Kouroupetroglou, G.: Modelling Emphatic Events from Non-Speech Aware Documents in Speech Based User Interfaces. In: 10th International Conference on Human-Computer Interaction, pp. 806–810 (2003)
12. Xydas, G., Argyropoulos, V., Karakosta, Th., Kouroupetroglou, G.: An Experimental Approach in Recognizing Synthesized Auditory Components in a Non-Visual Interaction with Documents, In: 11th Int. Conference on Human-Computer Interaction (2005)

13. Gussenhoven, C.: Intonation and interpretation: phonetics and phonology. In: Speech Prosody 2002, pp. 47–57, URL http://aune.lpl.univ-aix.fr/sp2002/pdf/gussenhoven.pdf (2002)
14. Launay, J., Segalen, L., Kanellos, I., Moudenc, T., Otesteanu, C., David, A., Fang, G., Jin, J.: Speech expressiveness: Modeling and implementing the expressive impact of typographic and punctuation marks for textual inputs. In: 3rd International Conference on Information and Communication Technologies: From Theory to Applications, pp. 1-6 (2008)
15. ACSS. Aural Style Sheets: http://www.w3.org/TR/CSS2/aural.html
16. Kallinen, K., Laarni, J., Ravaja, N., Saari, T.: Auditive "Boldfacing", Emotional Characteristics of Message, and Individual Differences in Memory Acquisition of Computer Mediated Business News. In: 11th Conference on Human-Computer Interaction (2005)
17. Truillet, P., Oriola, B., Nespoulous, J.L., Vigoroux, N.: Effect of Sound Fonts in an Aural Presentation. In: 6th ERCIM Workshop, UI4ALL, pp. 135-144 (2000)
18. Argyropoulos, V., Sideridis, G., Kouroupetroglou, G., Xydas G.: Auditory Discriminations of Typographic Attributes of Documents by Students with Blidness, British Journal of Visual Impairment, 27(3), pp. 183-203 (2009)
19. Raman, T.V.: An Audio view of (LA)TEX Documents. In: Annual Meeting. TUGboat, 13(3), pp. 65-70 (1992)
20. Kouroupetroglou G., Tsonos, D.: Multimodal Accessibility of Documents. Chapter in the book: Shane Pinder (ed.) Advances in Human-Computer Interaction, I-Tech Education and Publishing, Vienna, pp. 451-470 (2008)
21. Lang, P. J., Bradley, M., Culthbert, B.: International Affective Picture System (IAPS): Instruction Manual and Affective Ratings, Technical Report A-6, The Center for Research in Psychophysiology, University of Florida, U.S.A. (2005)
22. Kouroupetroglou, G., Tsonos, D., Vlahos, E.: DocEmoX: A System for the Typography-Derived Emotional Annotation of Documents. LNCS, vol. 5616, pp. 550-558 (2009)
23. Schröder M.: Expressing degree of activation in synthetic speech. IEEE Transactions on Audio, Speech and Language Processing, 14(4), pp. 1128-1136 (2006)
24. Tatham, M., Morton, K.: Expression in Speech: Analysis and Synthesis Oxford linguistics, Oxford University Press (2006)
25. Tsonos, D., Ikospentaki, K., Kouroupetrolgou, G.: Towards Modeling of Readers' Emotional State Response for the Automated Annotation of Documents. In: IEEE World Congress on Computational Intelligence, pp. 3252-3259 (2008)
26. Schröder , M., Trouvain, J.: The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology, 6, pp. 365-377 (2003)
27. Tsonos, D., Kaccori, H., Kouroupetroglou, G.: A Design-for-All Approach Towards Multimodal Accessibility of Mathematics. In: 10th International Conference of the Association for the Advancement of Assistive Technology in Europe (2009)
28. Tsonos D., Kouroupetroglou, G.: Accessibility of Board and Presentations in the classroom: a Design-for-All Approach. In International Conference on Telehealth and Assistive Technologies, pp. 13-18 (2008)
29. Ikospentaki, K., Vosniadou, S., Tsonos, D., Kouroupetroglou, G.: HOMER: A Design for the Development of Acoustical-Haptic Representations of Document Meta-Data for Use by Persons with Vision Loss. In: 2nd European Cognitive Science, vol. 25, pp. 912 (2007)
30. Tsonos, D., Kouroupetroglou, G.: Prosodic Mapping of Typographic Alterations based on the Dimensional Theory of Emotions. IEEE Transactions on Affective Computing, under review (2012)

31. Ikospentaki, K., Tsonos, D., Vosniadou, S., Kouroupetroglou, G.: Sonification of Board Typographic Elements: Does it Enchance the Traditional Teaching Approach of Sighted and Blind Students?. Journal of Visual Impairment and Blindness, in preparation (2012)
32. Katsoulis, F., Tsonos, D., Kouroupetroglou, G.: Analysis of Text Signaling Devices. International Journal on Document Analysis and Recognition, in preparation (2012)
33. Tsonos, D., Kouroupetroglou, G.: A Methodology for the Extraction of Reader's Emotional State Triggered from Text Typography. Chapter in the book: Paula Fritzsche (ed.) Tools in Artificial Intelligence, In-Tech Publishing, Vienna, pp. 439-454 (2008)
34. Tsonos, D., Xydas, G., Kouroupetroglou, G.: Auditory Accessibility of Metadata in Books: A Design for All Approach. LNCS, vol. 4556, pp. 436-445 (2007)
35. Tsonos, D., Xydas, G., Kouroupetroglou, G.: A Methodology for Reader's Emotional State Extraction to Augment Expressions in Speech Synthesis. In: 19th IEEE Int. Conf. on Tools with Artificial Intelligence, vol. II, pp. 218-225 (2007)