Experimentation on Spoken Format of Tables in Auditory User Interfaces

Dimitris Spiliotopoulos[1], Gerasimos Xydas[1], Georgios Kouroupetroglou[1] and Vasilios Argyropoulos[2]

[1]University of Athens, Department of Informatics and Telecommunications
Panepistimiopolis, Ilisia, GR-15784, Athens, Greece
{dspiliot, gxydas, koupe}@di.uoa.gr

[2]University of Thessaly, Department of Special Education
vassargi@sed.uth.gr

Abstract

The acoustic representation of complex visual structures involves both synthesized speech and non-speech audio signals. Though progress in speech synthesis allows the consistent control of an abundance of parameters, like prosody through appropriate mark-up, there is not enough experimentally proven specification input data to drive a Voice Browser for such purposes. This paper reports on the results from a series of psychoacoustic experiments aiming to provide natural speech prosodic specification for the task of vocalizing tables. Blind and sighted listeners were asked to reconstruct simple and complex data tables from naturally spoken descriptions. From the listeners' feedback it was deducted that consistent prosodic rendering can model the underlying semantic structure of tables.

1 Introduction

The vast majority of documents read by humans today are accommodated by visually oriented meta-information in the form of particular visual components. These range from simple, "bold", "italic", or coloured letters directives to more complex ones such as those that define a spatial layout (tables, forms, etc.). This work is concerned with the accessibility of complex document visual structures for both the visually capable and the visually impaired in cases of speech-only user interfaces.

A plain text document carries linear textual information that can be rendered to speech straightforwardly by screen reading applications or other types of software that read web pages and documents. On the other side, complex visual document structures, that yield non-linear information, are not efficiently presented in speech-only user interfaces. Most common approaches tend to linearize these two-dimensional elements prior to their acoustic presentation. However, most of the semantic meaning of their enclosed text is implicit to the visual structure.

Table is a generally used term to denote certain structural layout. This work is concerned with *data tables*, the most widely used two-dimensional structure in documents. These complex visual structures bear a distinct association between the physical layout and the underlying logical structure (Ramel, Crucianou, Vincent & Faure, 2003). Previous works show that appropriate markup can be used to assign logical structure to table cells (Hurst & Douglas, 1997) and suggest additional markup annotation to existing tables for adding context in order to improve navigation (Filepp, Challenger, & Rosu, 2002). Other suggestions include automated approaches for retrieval of hierarchical data from HTML tables (Lim & Ng, 1999).

Rendering tables in audio constitutes a hard task because of the difficulty in accessing the semantic information under the visual structure. Smart browsers are used to access critical information for use in reading tables as well as linearization techniques are employed for transforming tables into more a easily readable from screen readers form (Yesilada, Stevens, Goble & Hussein, 2004).

Previous works show that information about the semantic structure of HTML tables can be used to aid navigation and browsing of such visual components (Pontelli, Gillan, Xiong, Saad, Gupta & Karshmer, 2002). The important meta-information hidden in tables is reconstructed in order to provide a means for readers to comprehend the representation of tables. However, since the problem is addressed on the visual level, the major handicap of the linearized transformation approach to the actual spoken form remains. Moreover, other studies focusing on the speech representation show that one-dimensional elements such as bold and italic letters, gain their acoustic representation by the use of prosody control (Xydas, Spiliotopoulos & Kouroupetroglou, 2003). Other works deal with the acoustic representation of linear visual components using synthesized speech (Raman, 1992), while in (Xydas & Kouroupetroglou, 2001) we had presented a script-based open platform for rendering meta-information to speech using a combination of prosody modifications and non-speech audio sounds.

The exploitation of synthetic speech prosody parameterization necessitates the utilization of the human natural spoken rendition for tables. Our motivation is to examine the speech specification of table meta-information by analysing such natural spoken paradigms from human readers. This work reports on the psychoacoustic experimentation on spoken format of data tables utilising prosodic attributes. The experiments were set up in two stages. First, adept readers were asked to read out simple and complex tables using both free and guided speaking style. The second stage involved two groups of five blind and five sighted first-time listeners that were asked to listen to the spoken rendered tables and reconstruct them back to written form. Finally, using the acquired analysed speech data from the most preferred spoken renditions, we are concluding to a pilot specification concerning prosodic boundary tones and pause breaks describing header data cell relation and end of row interpretation.

2 Table Specification

Tables are used in abundance in electronic documents today. There exist several document type specifications that contain table descriptions allowing implementation of tables. One of the most common document types, used by the on-line community on an every day basis, is HTML, which provides meta-information about the way that text data is visually presented. According to the W3C, "The HTML table model allows authors to arrange data into rows and columns of cells", (Raggett, Le Hors & Jacobs, 1999).

We are interested in the vocalisation of table structures in auditory user interfaces. Transferring a structure from the visual modality to aural is but no means an easy task. Tables are characterized by many qualitative and quantitative aspects that should be taken into consideration since successful vocalisation is greatly affected by them. The available structures that can be characterised as "tables" can be of great diversity, the following factors be taken into account:

- Size of table
- Design of table
- Amount of data in each cell
- Type of data in each cell
- Visual enhancements
- Complexity
- Inherent underlying semantic structure
- Browsing

Considering the above points of interest, the *size* of a table can range from just a couple of cells (e.g. 1x2 table size) to any number of rows and columns. Table *design* refers to the implementation of the spatial layout by the designer. There is more than one way of successfully presenting the same information using different tables. Those tables in effect should convey the data in the way intended by the designer. The designer may use different header cells or grouping of data. The differences in design may be in the use or non-use of certain attributes. All sorts of data can be supported inside table cells. A table cell may contain any *amount* of data. (e.g. a whole sentence against one letter). The *type* of data may also vary. The data in HTML table cells, for instance, may be text, preformatted text, images, links, forms, form fields, other tables, and so on (Raggett et al., 1999). *Visual enhancements*, such as colour or bold letters, are commonly used by designers in order to emphasise certain parts (cells, rows, etc.) of a table. Tables may also vary in *complexity*, nested tables and cell merging being the most common examples. Last but certainly not least a factor is the underlying *semantic structure* of a table, for most people the most difficult parameter to utilise. *Browsing* is not a table characteristic but rather a manner of navigation and/or linearization of a table discussed in detail later (paragraph 2.4).

For the experiment in this work the focus is on the realisation of visual table structures to voice. To achieve that, certain care had to be taken to ensure that the tables used in the experiment were designed to be processed aurally, as well as conforming to certain guidelines for wellformedness. In respect to that, certain considerations pertaining to the above discussion were made.

2.1 Table accessibility

Special recommendations to promote accessibility, containing guidelines on how to make the web content accessible to people with disabilities are provided by the W3C (Chisholm, Vanderheiden & Jacobs, 1999). According to these, the use of `<TH>` (for headers) and `<TD>` (for data cells) is mandatory. The use of `<THEAD>`, `<TFOOT>`, and `<TBODY>` to group rows and `<COL>` and `<COLGROUP>` to group columns is also required to associate data and header cells. Moreover, W3C table description states that HTML data tables should not be used for visual layout purposes, that being the task of style sheets.

This work is concerned with model data tables which by design are compliant to high priority Web Accessibility Initiative (WAI) guidelines recommendation. The term *data tables*, from this point onward, will refer to HTML *TABLE* containers used solely to convey information comprising of pure data of certain relationship, not used for page layout and without any visual enhancements or styles applied. Such tables use HTML markup reserved for data tables, like <TH>, <TD>, <CAPTION>, etc. These elements are used by agents in order to identify and manipulate the data before vocalisation.

2.2 Underlying Semantics and browsing considerations

Tables are visual structures spanning two dimensions, used as a means of grouping information. The designer of a table is concerned with the visual representation of the data. The resulting table is a visual structure that was formed to accommodate the semantic relationship of the data. In fact, the semantic relationship is the cause for choosing a table as the most appropriate visual structure for modelling the information. The table as a structure models the semantic information visually. The semantic information is then inherited inside the structure in the visual modality and ideally can be retrieved by the visual reader. This is possible because the reader can see the whole structure and deduct the semantic relationships between the header and data cells by deciphering the visual representation. There is direct connection (Figure 1) between the original data (natural language), the structural information (written language), and the data reconstruction (natural language).

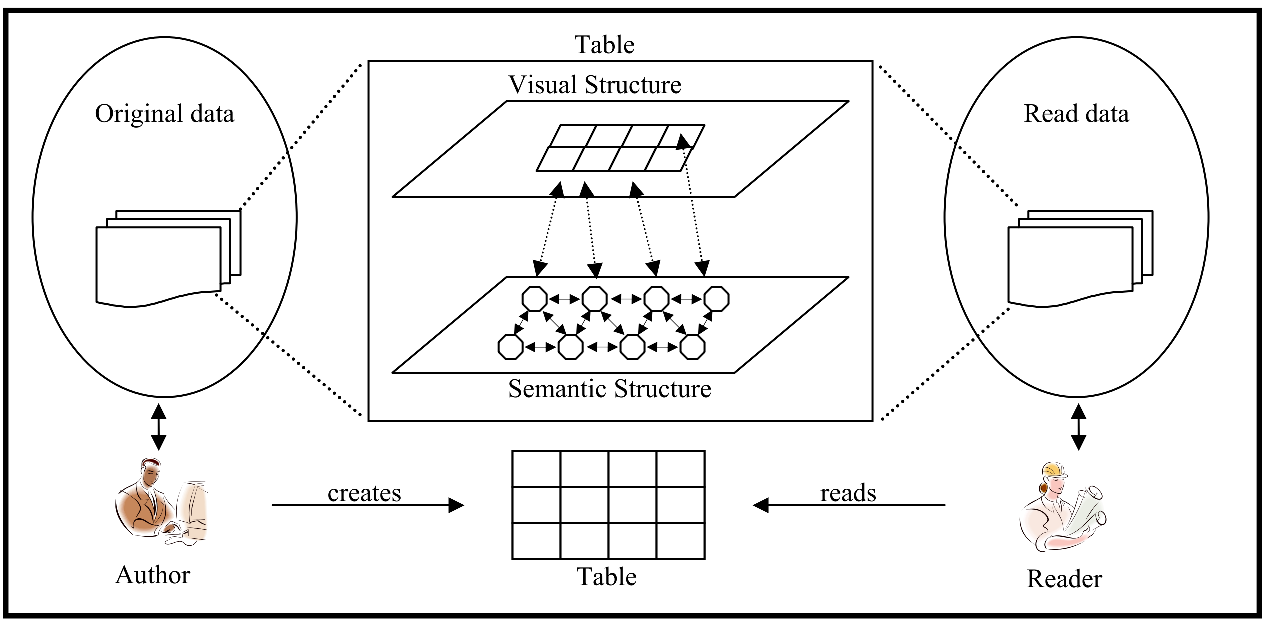

Figure 1: Table Structural Information – Creating and Reading a Table

Accessing table structures by visually impaired or blind people is a far more difficult task since the visual structure has to be processed in order to render it to aural modality. Auditory user interfaces access the tables and try to present the information to the listeners. There are two major considerations in this case. The first is the non-visual browsing of tables, which involves certain techniques in identifying and using appropriate linearization for aural rendition. The second is the acoustic representation of the linearized table that should be rich enough to convey the linearized table into understandable and faithful to the intended semantic structure speech (Figure 2).

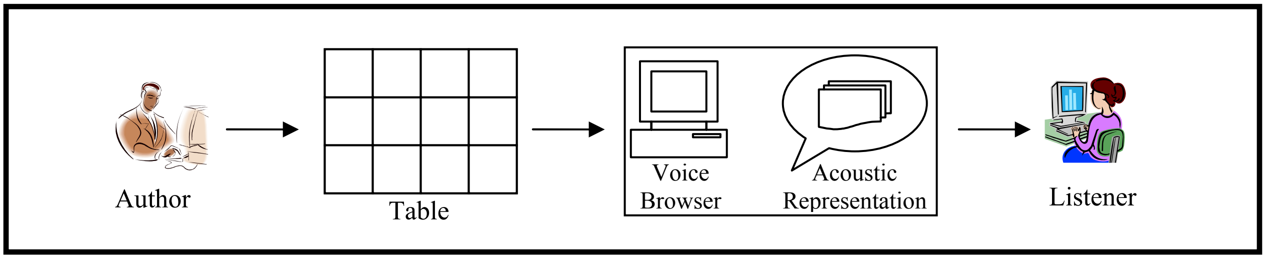

Figure 2: Spoken Format Rendering of Tables

Tables are visual structures that have no aural equivalent, thus the spoken form has to be constructed. This can be achieved by producing a linear form of the data contained in the table structure. The table can be rendered to speech in several ways depending on the approach selected for linearization. The common screen reader would linearize a table row by row, resulting in a loss of semantic information between the cell data. Figure 3 shows how linear reading of the example table retains no structural meaning, rendering the output almost invaluable to the listener.

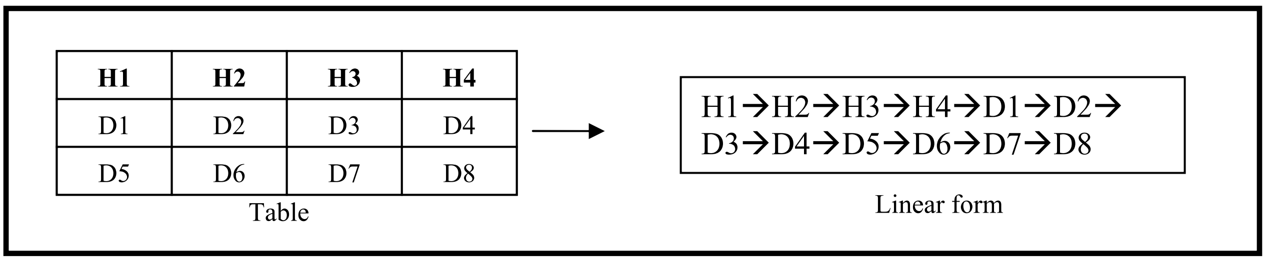

Figure 3: Common Linear Reading of Table Data

Recent research shows that advanced browsing techniques may be used to create table linearization that analyses the implicit structural information contained in tables so that it is conveyed to text (and consequently to the listeners) by navigation of the data cells (Pontelli, Xiong, Gupta & Karshmer, 2000). Figure 4 depicts examples of the same table rendered in two different forms according to specification.

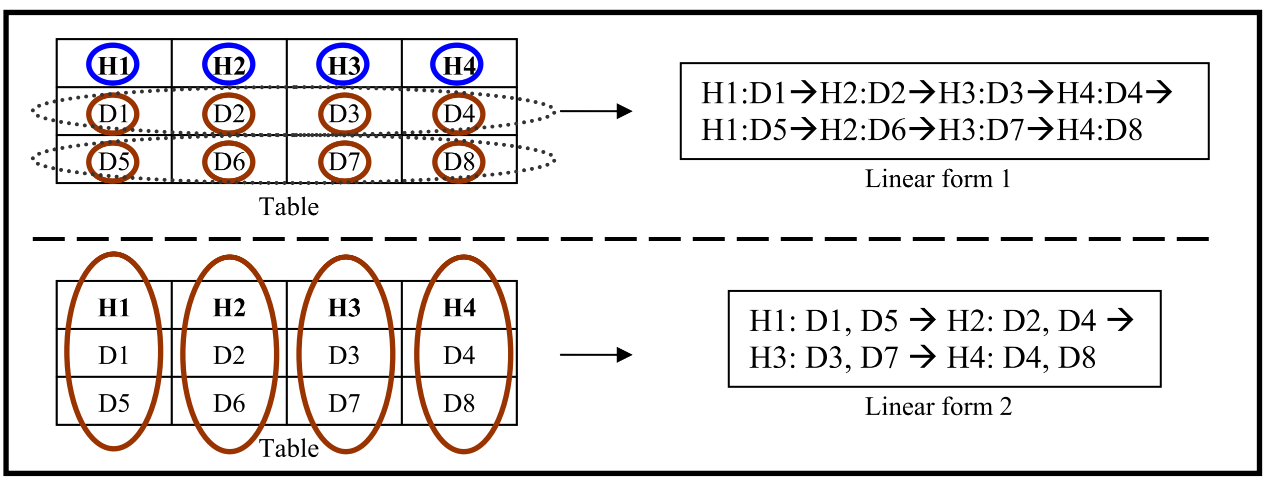

Figure 4: More Sophisticated Linear Readings of Table Data

2.3 Table Complexity

Completing the previous consideration, emphasis is given to the use of wellformed data tables taking into account the variation in complexity that these structures may carry. As mentioned previously, tables are structures that are used to arrange content into a two-dimensional yet semantically coherent assembly. Cells are the basic blocks that contain data. The type of data may be “header” information, which is used to describe the data in other cells, or “data” information. Cells are arranged into rows and columns that are perceived as groups of data labelled by the header information, thus forming the table.

In terms of complexity, data tables can be categorized into two types, simple and complex. Simple data tables have up to one row and up to one column of header cells. Figure 5 shows an example of a simple table used in the literature. It is easy to observe that the first row is the only row of header cells while there is no respective column of headers.

Name	Phone Number	Age	Weight
Steve Nelson	425/555-2186	54	130 lbs.
Maria Sanchez	425/555-8741	43	120 lbs.

Header row

Data rows

Figure 5: A Simple Data Table Example

Complex data tables contain more than one level of logical row or column headers. This means that header cells can be expanded to encompass more than one row or column. Moreover, the same can be true for data cells. This can lead to whole tables existing as nested tables forming a complex table. From the recognizable example complex table shown in Figure 6, it is obvious that complex data tables are much more complicated to browse since they may have multiple levels of structures in a hierarchical manner.

Travel Expense Report				
	Meals	**Hotels**	**Transport**	**subtotals**
San Jose				
25-Aug-97	37.74	112.00	45.00	
26-Aug-97	27.28	112.00	45.00	
subtotals	65.02	224.00	90.00	379.02
Seattle				
27-Aug-97	96.25	109.00	36.00	
28-Aug-97	35.00	109.00	36.00	
subtotals	131.25	218.00	72.00	421.25
Totals	196.27	442.00	162.00	800.27

<CAPTION>

Headers

Simple tables

Figure 6: A Complex Data Table Example

The example complex table can be thought of as a three-dimensional structure (Pontelli et al., 2000), compared to the two-dimensional simple data table of Figure 5. Figure 7 shows a semantic structure dimensional comparison view of the two tables. The third dimension of the semantic structure of the complex table is embedded inside the two dimensional visual structure.

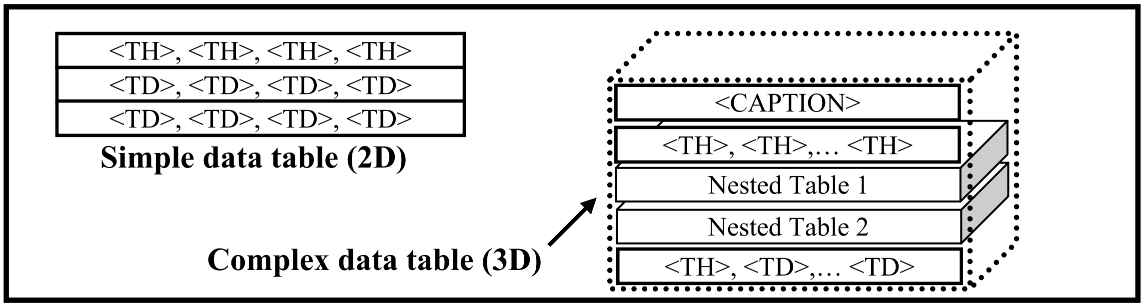

Figure 7: Simple and Complex Data Table Dimensional Comparison

3 An Experiment in Reading Tables

From the above considerations it seems prudent to select proper *data tables* for the experiment – the term *genuine tables* is sometimes used as a definition for tables where the two dimensional grid is semantically significant (Penn, Hu, Luo & McDonald, 2001). Both simple and complex data tables have to be considered in order to observe data semantic coherence to gain enough knowledge for representing such visual structures acoustically. The semantic representation must be consistent and straightforward for the readers and the listeners to accurately fathom. The examined tables have to be well-formed data tables conforming to W3C WAI Guidelines.

The aim of this psychoacoustic experiment was the examination of speech prosody characteristics that describe the human aural rendition of data tables. Natural spoken rendition required human subjects as readers of data tables. The tables that were selected for this experiment were similar to the reference simple and complex tables described earlier in Figures 5 and 6. The data in both tables was in Greek, the native language of all readers and listeners. Tables 1 and 2 show the exact simple and complex data tables used in the experiment.

Table 1: The Simple Data Table Used in the Experiment

Πληροφορίες πελατών[1]

Όνομα	**Τηλέφωνο**	**Ηλικία**	**Περιοχή**
Θωμάς Μίζας	4850918	54	Ιλίσια
Μαρία Ρίπη	7776755	29	Μαρούσι
Έλενα Ρήγα	5640089	43	Παγκράτι
Νίκος Μίστας	3867244	32	Θησείο

Table 2: The Complex Data Table Used in the Experiment

Αναφορά δαπανών ταξιδίου[1]

	Γεύματα	**Διαμονή**	**Μεταφορά**	**Σύνολο**
Παρίσι				
25-Αυγ-97	37.00	112.00	31.00	
26-Αυγ-97	28.00	112.00	31.00	
Μερικό σύνολο	65.00	224.00	62.00	351.00
Λονδίνο				
27-Αυγ-97	96.00	109.00	16.00	
28-Αυγ-97	35.00	109.00	16.00	
Μερικό σύνολο	131.00	218.00	32.00	381.00
Γενικό σύνολο	196.00	442.00	94.00	732.00

[1] This is the table caption.

The psychoacoustic experiments were set up in two stages. First, three adept readers - each an expert in their relevant fields: linguistics, speech synthesis, and musicology/acoustics - were asked to read out (a) the simple and (b) the complex tables. The readers were fluent individuals bearing adequate familiarization with table structures and prosody. They were asked to provide two spoken formats for every table, following the suggestions for table browsing that were discussed earlier. One reading was made using an almost free style similar to the plain linear browsing, and the other was a guided navigation matching the "Linear form 1" as illustrated in Figure 4. The readers were encouraged to provide their own prosodic rendition of the tables in accordance with their subjective judgment and expertise.

The second stage involved two groups of five blind and five sighted first-time listeners that were asked to listen to all acoustic representations of the tables in random order and reconstruct them back to written form (a) on paper (sighted subjects), or (b) on Braille (blind subjects). Before starting the experiment, the listeners (five male and five female, aged 21-24 years old) were shown sample data tables (printed matter for the sighted, Braille for the blind) of simple and complex tables and were given ample time to familiarize themselves with the examples.

At the start of the experiment the participants were asked to provide a brief definition of the data table. The objective task was to reconstruct the table information the best possible way bearing in mind that the implicit structure should be retained in order to be allowed to be accessed from the written data in the best possible way. The following experiment setup was followed:

- For each table rendition style (freestyle and guided for simple and complex tables):
    - For each table reading, the listeners assigned:
        - Confidence factor, 1 (lowest) -5 (highest)
        - Acceptance factor, subjective evaluation score, 1 (lowest) -5 (highest)
        - Brief justification and/or problem indication for subjective evaluation score
    - At end of each style: subjective preference of reader rendition (choice of acoustic rendition)

4 Results

The experimental results were evaluated in order to classify table reconstructions from the subjects. The confidence score from all listeners was 4 or above from scale 1 (lowest) to 5 (highest) which is enough to validate all scoring. The ES evaluation score range of 1 (lowest) to 5 (highest) was distributed in respect to the following considerations:

- 1 - non reconstructive 2-D pattern, non comprehensible, vast lack of data,
- 2 - existence of a vague reconstructive 2-D pattern, non comprehensible, a great amount of data is missing,
- 3 - existence of a clear reconstructive 2-D pattern, quite comprehensible but still an amount of data is missing,
- 4 - existence of a clear 2-D reconstructive pattern, comprehensible with coherence, few data is missing
- 5 - perfect representation of a 2-D pattern, comprehensible with coherence, all data is included in the representational form

Table 3 shows the LSES (listener subjective evaluation score) and EOES (examiner objective evaluation score) score averages from blind and sighted subjects for all four data table natural speech renditions from all readers.

Table 3: Partial and Overall Scores for Simple and Complex Structure Reconstruction and User Feedback

	Simple (freestyle)		**Simple (guided)**		**Complex (freestyle)**		**Complex (guided)**	
	LSES	**EOES**	**LSES**	**EOES**	**LSES**	**EOES**	**LSES**	**EOES**
Blind	3.5	3.8	3.6	3.9	1.7	1.3	3.9	3.2
Sighted	3.1	4.9	4.4	4.8	2.7	2.7	3.9	4.7
Overall	3.3	4.4	4.0	4.4	2.2	2.0	3.9	4.0

Initial examination of the results shows that common linearization of the complex data table resulted in poor comprehension from all listeners with the average score of 2.0. On the other hand, the average score for the respective rendition of the complex data table following the proposed browsing of such structures was 4.0, confirms the proposed browsing from research sited earlier. Further observation shows that the sighted subjects performed better, something that was actually expected. Blind subjects scored about 1 mark lower than sighted subjects in the simple table cases while about 1.5 lower in the complex table ones. That is a very consistent behaviour, and possible indication of a pattern that may be useful in further similar experiments.

Amongst the three spoken renditions (for each of the four table-style combinations), all listeners (and the respective table reconstruction evaluation) seemed to agree on specific renditions for the complex tables, while disagreeing on the simple tables cases. Table 4 shows the spoken rendition preferences. Nevertheless, consistency of the results was preserved between reading styles.

Table 4: Natural Speech Rendition Preference (SR: spoken rendition)

	Simple (freestyle)	**Simple (guided)**	**Complex (freestyle)**	**Complex (guided)**
Blind	SR-2	SR-2	SR-2	SR-2
Sighted	SR-1	SR-1	SR-2	SR-2
Overall	SR-1+ SR-2	SR-1+ SR-2	SR-2	SR-2

From this point onwards the "SR-X" will refer to the spoken rendition from a reader for a particular table-style combination, where "X" is the reader number.

The spoken renditions made by reader 3 were not clear preferences for any of the table and reading style, although they ware highly valued among the listeners. The reason is that the other two readers' renditions offered a combination of characteristics that made them better performers. Moreover SR-3s were judged to be a bit on the faster side in terms of speech rate as well as in pause breaks. As a note for the future, from the respective subjective comments, the SR-3s could be thought of as a lower threshold in speech rate and pause breaks for reading tables or similar structures (or perhaps, as suggested by some listeners, could form a basis for parameterization for use when repeating a structure).

The analysis of speech signals of the spoken renditions themselves revealed several qualitative differences. It involves prosody analysis of the signals in terms of phrase accent tones, pause breaks, as well as overall speech rate. For prosody markup, the ToBI annotation model (Silverman et al., 1992) was used as a means of qualitative analysis. In this experiment a key factor is the realisation that the structural layout of the tables governs the speech rendition. Thus, the analysis was focused specifically on boundary tones (L- describing the fall in pitch at end of spoken data cell, H- describing the boundary rise) and respective pause breaks between cells and rows, for each of the four spoken renditions. Table 5 depicts the placement of accent tones for SR-1 and SR-2. In this case most listeners enjoyed the change in pitch on the penultimate cell of the <TD> rows, a clear indication of the end of row.

Table 5: Simple Table (freestyle) Boundary Tones

	SR-1	**SR-2**
<CAPTION>	L-	L-
<TH> Rows	H- H- H- L-	H- H- H- L-
<TD> Rows	L- L- **L-** L-	L- L- **H-** L-

Table 6: Simple & Complex Table (guided) Boundary Tones

	SR-1	**SR-2**
<CAPTION>	L-	L-
<TH>	H-	H-
<TD>	L-	L-

Table 6 shows that both guided spoken renditions for the simple table shared the same characteristics, fully approved by the listeners. All <TH> cells had high phrase accent and all <TD> cells low. Similar results were obtained for complex table SRs. As shown in Table 7, a point of high impact was observed for the freestyle linear renditions. The change of pitch when reading the 5^{th} and 9^{th} rows (the last row of each nested data table - worked out as indication of end of block of data) clearly resulted in better reconstruction. Both readers tried to acoustically represent the existence of nested tables, clearly rendering the headers of the nested tables as L- (first cell data in 3^{rd}-4^{th} and 7^{th}-8^{th} rows) as well as the end of those tables (SR-2). However the results were far from acceptable as explained earlier.

Table 7: Complex table (freestyle) accent tones

	SR-1	**SR-2**
<CAPTION>	L-	L-
1st Row	H- H- H- L-	H- H- H- L-
2nd Row	L-	L-
3rd Row	L- H- H- L-	L- H- H- L-
4th Row	L- H- H- L-	L- H- H- L-
5th Row	**L-** H- H- H- L-	**H-** H- H- H- L-
6th Row	L-	L-
7th Row	L- H- H- L-	L- H- H- L-
8th Row	L- H- H- L-	L- H- H- L-
9th Row	**L-** H- H- H- L-	**H-** H- H- H- L-
10th Row	L- H- H- H- L-	L- H- H- H- L-

The pitch analysis for the guided SRs for the complex table showed identical results and the same pattern as for the simple table (see Table 6).

Pause breaks were expected to play a significant role in the spoken format of tables since pausing could prove to be very helpful in distinguishing between rows. That is true for data tables because rows are used by the designer to represent semantic correlation between the data they encompass. Besides, pause breaks would prove to be most useful in cases where the pitch remains unchanged. Such example is shown in tables 8 and 9.

Table 8: Simple table (freestyle) breaks (in seconds)

	SR-1	**SR-2**
End of cell	2.31	1.09
End of row	2.91	1.75

Table 9: Simple table (guided) breaks (in seconds)

	SR-1	**SR-2**
End of header cell	0.60	0.82
End of data cell	2.41	2.10

For simple tables (freestyle), the pause at the end of each cell data and end of row of cell data proved more significant than change in pitch for the sighted listeners. The blind listeners were more sensitive to pitch, as shown by other works (Xydas, Argyropoulos, Karakosta & Kouroupetroglou, 2005). The guided renditions were affected also. The difference in pause breaks made the header-data pairs more distinct for the listeners. Shorter break between the two pieces of data combined with much longer pause between pairs resulted in better scoring. In this case the blind subjects achieved better results from SR-2 since the notion of pairs was deprecated by shorter pause after header cell data.

For complex tables, pitch variation had a greater impact than pause length. In the first case (freestyle), there was little different in pause breaks, as shown in Table 10. The pitch variation made all the difference, although the linear rendition of the structure left only little to be salvaged. The situation was different for the guided renditions (Table 11). This time, semantics were carried over to the listeners and allowed prosody - in this case the pause length - to outline the structure. Longer pauses in SR-2 (ranging from 49-64% longer than SR-1), and enough variation between header-data cell pairs and “end of row” marking provided the most successful vocalization.

Table 10: Complex table (freestyle) breaks (in seconds)

	SR-1	**SR-2**
End of header row	2.41	2.55
End of data row	2.80	2.66
Top row header cell	1.42	0.92
Other header cell	2.34	2.00
Data cell	1.69	1.74

Table 11: Complex table (guided) breaks (in seconds)

	SR-1	**SR-2**
Nested table header	1.84	2.79
End of row	2.32	3.81
Header cell	1.10	0.68
Data cell	1.94	2.83

Speech rate, in this experiment was a factor with minimal effect since there was no connected speech involved. In fact, the strong presence of long pause breaks renders any speculation worthless.

5 Conclusion

We presented an experimental study of spoken presentation of data tables. Human readers were employed to render both simple and complex cases evaluated by blind and sighted listeners. A set of prosodic parameters was analyzed in terms of boundary tones and pauses, clearly illustrating consistency against cell content and visual structure. The deducted specification can form the basis for auditory scripting modeling of tables, to aid automatic rendering using synthetic speech.

6 Acknowledgements

The work described in this paper has been partially supported by the HERACLITUS project of the Operational Programme for Education and Initial Vocational Training (EPEAEK) of the Greek Ministry of Education under the 3rd European Community Support Framework for Greece. A special thanks to all the students that have taken part in the psychoacoustic experiments described in this work.

References

Chisholm, W., Vanderheiden, G., & Jacobs, I. (1999). Web Content Accessibility Guidelines 1.0, W3C Recommendation, 5 May 1999, *http://www.w3.org/TR/WAI-WEBCONTENT/*

Filepp, R., Challenger, J., & Rosu, D. (2002). Improving the Accessibility of Aurally Rendered HTML Tables. *Proc. ACM Conf. on Assistive Technologies - ASSETS 2002*, 9-16.

Hurst, M., & Douglas, S. (1997). Layout & Language: Preliminary Experiments in Assigning Logical Structure to Table Cells. *Proc. 4th Int. Conf. Document Analysis and Recognition – ICDAR 2003*, 1043-1047.

Lim, S., & Ng, Y. (1999). An Automated Approach for Retrieving Hierarchical Data from HTML Tables. *Proc. 8th ACM Int. Conf. Information and Knowledge Management - CIKM 1999*, 466-474.

Penn, G., Hu, J., Luo, H., & McDonald, R. (2001) Flexible Web Document Analysis for Delivery to Narrow-Bandwidth Devices, *Proc. 6th Int. Conf. on Document Analysis and Recognition - ICDAR 2001*, 1074-1078.

Pontelli, E., Gillan, D., Xiong, W., Saad, E., Gupta, G., & Karshmer, A. (2002). Navigation of HTML Tables, Frames, and XML Fragments. *Proc. ACM Conf. on Assistive Technologies - ASSETS 2002*, 25-32.

Pontelli, E., Xiong, W., Gupta, G., & Karshmer, A. (2000). A Domain Specific Language Framework for Non-visual Browsing of Complex HTML Structures. *Proc. ACM Conf. Assistive Technologies - ASSETS 2000*, 180-187.

Raggett, D., Le Hors, A., & Jacobs, I. (1999). Tables, HTML 4.01 Specification, W3C Recommendation, *http://www.w3.org/TR/REC-html40*

Raman, T. (1992). An Audio View of (LA)TEX Documents, *TUGboat, Proc. 1992 Annual Meeting*, 13, 3, 372-379.

Ramel, J-Y., Crucianou M., Vincent, N., & Faure, C. (2003). Detection, Extraction and Representation of Tables. Proc. *7th Int. Conf. Document Analysis and Recognition - ICDAR 2003.* 374-378.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A Standard for Labeling English Prosody. *Proc. Int. Conf. Spoken Language Processing - ICSLP-92*, 2, 867-870.

Xydas, G., & Kouroupetrolgou, G. (2001). Text-to-Speech Scripting Interface for Appropriate Vocalisation of E-Texts. *Proc. 7th European Conf. Speech Communication and Technology - EUROSPEECH 2001*, 2247-2250.

Xydas, G., Argyropoulos, V., Karakosta, T., & Kouroupetroglou, G. (2005). An Experimental Approach in Recognizing Synthesized Auditory Components in a Non-Visual Interaction with Documents. *Proc. Human-Computer Interaction - HCII 2005*, to appear.

Xydas, G., Spiliotopoulos D., & Kouroupetroglou, G. (2003): Modeling Emphatic Events from Non-Speech Aware Documents in Speech Based User Interfaces. *Proc. Human-Computer Interaction - HCII 2003, Theory and Practice*, 2, 806-810.

Yesilada, Y., Stevens, R., Goble, C., & Hussein, S. (2004). Rendering Tables in Audio: The Interaction of Structure and Reading Styles. *Proc. ACM Conf. Assistive Technologies - ASSETS 2004*, 16-23.