

Understanding Genesis of Cyber Signs and Their States - Designed for the Web - An Example: Development of Speech Synthesis Markup Language

Authors: Zlatko Papes, Davorin Bengez, Vesna Kirinic Papes

Abstract

Understanding tendencies in the stratification of structural layers of "cyber" sign: client, server, stylistic, perceptive, presentation, application, data oriented. Signs for both, humans and machines. From signs which present distinctive perceptive forms to signs which present distinctive meaning with metadata identity - and back. From representational (HTML) to a semantic web with separated rendering (HTML+CSS). Toward pragmatic signs spaces in structuring the Web - application oriented, platform independent, open and safe. In less expected environment with applets and Java. In predefined situation in some way - with some sort of XML languages, on example, forms oriented - with FormsML, mobility oriented - WML, more visual intelligibility - VRML, and other aural or multimodal oriented signs. Frequent incidents of laptops being stolen versus centralized delegating of interactivity, logic filtering and content adaptation on standardized fixed and mobile web client interfaces. Modification of "access from everywhere, whereas maintenance of data integrity located in one place", in slightly new trends started with TopClass player or Oracle 8i Lite. Transformational limitations of XSLT and search for other solutions of dynamic web-content transformation with general scripting languages, such as Perl, Python, PHP, Tcl/TK or new XML-script. Understanding genesis of voice structured marked content (XML) in contrast to non-structured representation recorded raw - .au, .wav or other streaming content. Edinburgh Speech Tools, Mbrola, Festival and SABLE. Significance and examples of text-to-speech technologies on Croatian spoken language, Zagreb Phonetic School and tendencies of developing Speech Synthesis Markup Language.

Preface (About Pragmatic in Sign)

Somewhere between the hope and the scientific determination of possibilities, as a result of effort to accelerate communication, there is a world of signs which opens up possibilities. Such signs prepare - or design what is possible to expect. A precondition work before probabilistic calculation. Made by a creative, culturally dependent act. It is designed for a successful perception, designed as art by a man and machine. It was first intended for both, for humans and machines, with a man as the final instance. Authors who supply media with such signs are trying to lead our attention or to give us a time for a glimpse, or to concentrate our perceptive (affective) energy on new things. They are trying to transfer intention beyond text on ASCII.

Cyber Sign environment (In What Kind of Space Cyber-Signs and Cyber-Designs Grow)

Cyber signs grow in a special space - Internet space - which is comprised a network of servers and static or mobile clients on one side and humans being on the other. There is an entire "virtual space" growing in front of our eyes, with properties that are a combination of intellectual and technological achievements and human perceptive characteristics. After the first ASCII era of documents (with ASCII - still in kept in tag names), digitally based communication was accelerated by the development of pragmatic signs, signs environment, or languages for the Internet space. This has given us more explicit control over communication design and message presentation. one realization of pragmatic function of languages are *markup languages*. (*Markup is text added to the data of a document in order to convey the information on that document. Document is a logical construct that contains a document element just like a book, contains chapter, paragraph and picture elements.*) They are derived from *Standard Generalized Markup Language (SGML)* (<http://www.iso.ch/>) [1] which was designed for a *document structure description*. Their tags grow, first oriented only in presentation as one of possible procedures - how typography formats the text on screen (HTML) and second, with auto description in tag names - for *semantic Web*.

Semantic Web (with *eXtended Markup Language, XML*) will take more searchable and more differentiated information, says Tim Berners Lee [2]. /He "directs the [World Wide Web consortium \(http://www.w3.org/\)](http://www.w3.org/), an open forum of companies and organizations with the mission to lead the Web to its full potential. With a background of a system design in real-time communications and text processing software development in 1989, he invented the World Wide Web - an internet-based hypermedia for global information sharing - while working at CERN, the European Particle Physics Laboratory. He wrote the first web client (browser-editor) and server in 1990"./

Web-designers create such signs through an investigation of a technological progress and properties of human perceptual abilities. Cyber - or web-sign, became the quickest way for documented communication in the world. Their increase asks for specific conditions or "design criteria".

Due to W3C-u, we can take a look at "work in progress" by reading the documentation about genesis of cyber signs for *generating speech* and *text-to-speech* in document processing: **W3C Working Draft 08 August 2000.: Speech Synthesis Markup Language Specification for the Speech Interface Framework**. It became usable, but only after considerable effort had been invested in a development of specific, local adaptation and standardization. It's genesis has characteristics of all cyber signs, and is very instructive in resolving problems in linguistic and audio media. For Croatia, their importance lies in possibilities of developing Croatian computer generated speech, today.

Design Criteria for (TextToSpeech) Cyber-Sign (Emphasized in citation added by authors who cited)

Here, we will show the relevance and importance of several cyber sign's genesis criteria

?? **general**

- o "A Text-To-Speech (TTS) system that supports the Speech Synthesis Markup Language will be responsible for rendering a document as spoken output and for using the information contained in the markup to render the document as intended by the author."
"The markup language - Speech Synthesis Markup Language is designed to be sufficiently rich... so that the document author (human or machine) can control the final voice output." [3]

?? **non-proprietary - open**

- o using the best strategical partner - vendor independent, high quality products (<http://www.opencontent.org/openpub/>, <http://www.gnu.org/>), [4],[5],[6]

?? **generated automatically, by document author, or both**

- o "*Document creation*: A text document provided as input to the TTS system may be produced automatically, by human authoring or through a *combination* of these forms. The Speech Synthesis markup language defines the form of the document."

?? **portability**

- o **platform independent** speech synthesizer from different companies (IBM, Lernout & Hauspie, Festival, etc.)

?? **easy controllable with standard markup scheme intended to by synthesizer independent and without proprietary tags, based on XML/SGML a widely used standard (to what degree and with which tools):**

- o is **textual** /cut and paste/
- o is **structural /modular, eXtensible**, DocBook (<http://www.oasis-open.org/docbook/>), [XML-Spy](#), [XML-script](#)../
- o is **interoperable** with other Markup Languages (Dialog ML - DML, Synchronized Multimedia Integration Language - SMIL, Aural Cascading Style Shit - ACSS), **synchronized facial animation** ("nice to have" - status) [7]

?? **interactive**

- sensible on touch, events, state of record in data-base environment/
- ?? **with defined non-markup default behaviour**
- ?? **with standardized phoneme alphabets /localization, variants of localization and combination in same document/**
- ?? easy **searchable**
 - semantic oriented XML - not stream or bitmap
- ?? with predefined **voice** layer (male, female, old, young) and with predefined dialog forms - VoiceXML
- ?? based on **host** or **client (voice browser) environment** - with speech but also with audio inserts /stream, midi.../ capabilities
- ?? **with document type definition - DTD**
 - describe possible structures of elements: which elements and in what order as standard for interchange for the Speech Synthesis Markup Language - SSML
- ?? with **prosodic** layer independent from application program interfaces
- ?? based on **different** application program interfaces -API
 - Java speech API, Microsoft speech API - provide tags for speaker directive, "engines". Synthesizers: Bell Laboratories, Center for Speech Technology Research - Edinburgh University, Mbrola waveform synthesizer

Example of Speech Generation Technologies (based on work of Belgian Faculty Polytechnique de Mons and Zagreb Phonetic School)

<http://tcts.fpms.ac.be/synthesis/> [8] has made an high quality speech diphone *waveform synthesizer* named MBROLA - say as "umbrella" (supported by speech synthesizer Festival) and is trying to engage many nations for non commercial, non military applications of their multilingual speech synthesizer, freely available, with number of diphone database. Based on Mbrola suggestion, two authors Juraj Bakran and Nikolaj Lazic, at the University of Zagreb, together with the students of phonetics at the Department of phonetics at philosophical Faculty, University of Zagreb, have prepared Croatian diphone database, which everyone can download for non-profit purposes. [This database](#) may not be sold or incorporated into any product with commercial purposes without prior permission from the Diphone Database Owner (<mailto:juraj.bakran@ffzg.hr> and nikolaj.lazic@ffzg.hrsecr@babeltech.be). [9] After conversion in stream, you can download and listen an example of a certain quality without Mbrola synthesizer: [mbrola-pjeva.wav](#) , [nahrvatskom.wav](#), [na-hrvatskom-Sun.au](#), [zalostan.wav](#) ([10]).

On the basis of their long research in speech semantics, specially in affective stylistic, intonation, phoneme, and speech perception, their results in speech rehabilitation, foreign language perception, and education, Zagreb phonetic school insists on an understanding that meaning can be realized also in only one syllable, or voice or segment of intonation, which can be smaller than a word or longer then sentence or discourse. This approach challenges a classical linguistic segmentation in communicative meaning and pragmatic research.

Instead of Conclusion (States of Speech Cyber Sign Today)

In September 1997, three labs, **Sun** Microsystems, **Bell Labs**, and **Edinburgh** University 's Center for Speech Language Technology forces to merge two proposed standards, namely STML, and Sun's Java Speech Markup Language, JSML. The consortium, called SaBLE, aims to produce common standard for TTS markup within the next year or so. It must be system independent standard for making up text for the purpose of synthesis.

Specification evolved as an initiative to combine three existing speech synthesis markup languages: SSML, the Speech Synthesis Markup Language, STML, the Spoken Text Markup Language, JSML,

the Java Synthesis Markup Language with goals: SABLE enables markup of TTS text input, for improving the quality and appropriateness of speech output, multilinguality, easy of use, portability and extensibility.

Aural Cascading Style Sheets specifications should be implemented by a voice browser in a synthesizer independent fashion. Audio formatters consume XML content and ACSS to produce a SABLE stream suitable for sending to a TTS-engine.

HTML Document + ACSS --> Audio Browser which converts HTML text into SABLE using ACSS (automatic audio formatters) --> TTS System interprets SABLE text provided by Audio Browser.

?? Intonational Controls

- o "Intonation also reflects emotion and many less definable characteristics that were not planned to be included in this specification" [3]

?? - "low-level" elements: Fine-Grained Acoustic-Prosodic Control

- o "It is anticipated that low-level markup will be generated by automated tools, so that compactness will be given priority over readability." [3]

Cyber signs are still developing, but now we can help their genesis [11]. [12].

References

1. Informational processing - Text and office systems - Standard Generalized Markup Language (SGML), First edition 1986-10-15, International Standard ISO 8879 -1986 (E), UDC 681.3.06
2. Berners Lee, T: [Design Issues: Technical and philosophical notes on web architecture](http://www.w3.org/DesignIssues/) (<http://www.w3.org/DesignIssues/>, <http://www.w3.org/DesignIssues/Semantic.html>)
3. [Speech Synthesis Markup Language. Specification for the Speech Interface Framework](http://www.w3.org/TR/2000/WD-speech-synthesis-20000808/), W3C Working Draft 08 August 2000 (<http://www.w3.org/TR/2000/WD-speech-synthesis-20000808/>)
4. Kirinic-Papes, V Papes, Z: [Znacaj i mogucnosti stvaralacke uporabe informacijskih tehnika u drustvenim predmetima za strategiju informacijskog razvitka Hrvatske](http://salata.mef.hr/Solaris2000/Zlatko/), Savjetovanje HDPIO, Solaris 2000 (<http://salata.mef.hr/Solaris2000/Zlatko/>)
5. Bengez, D: [Sto moze otvoreni sistemski softver pomoci nasem skolstvu](http://salata.mef.hr/Solaris2000/davor-html/), Savjetovanje HDPIO, Solaris 2000 (<http://salata.mef.hr/Solaris2000/davor-html/>, <http://salata.mef.hr/Solaris2000/Davor/>)
6. Konjevoda, P: [Free software za analize podataka: matematika, statistika, data mining](http://salata.mef.hr/Solaris2000/Pasko/) (<http://salata.mef.hr/Solaris2000/Pasko/>)
7. [Center for Spoken Language Understanding](http://cslu.cse.ogi.edu/toolkit/), Oregon Graduate Institute of Science and Technology, (<http://cslu.cse.ogi.edu/toolkit/>)
8. [Belgian Faculte Polytechnique de Mons](http://tcts.fpms.ac.be/synthesis/) (<http://tcts.fpms.ac.be/synthesis/>)
9. J. Bakran, N. Lazic (1998) Fonetski problemi difonske sinteze hrvatskoga govora, Govor, XV, br.2, 103-116. : <http://tcts.fpms.ac.be/synthesis/mbrola/dba/cr1/cr1-981028.zip>
10. [The Programmer's file format collection](http://www.wotsit.org) <http://www.wotsit.org>
11. Papes, Z Bengez, D Marijanovic, M: [Labs for Knowledge Presentation](http://www.carnet.hr/cuc/cuc99/radovi/B1/b1-4f/index.html), CuC 1999 (<http://www.carnet.hr/cuc/cuc99/radovi/B1/b1-4f/index.html>)
12. Papes, Z, Kirinic-Papes V: [Okosnice infrastrukturne podrske informatickom obrazovanju u Hrvatskoj](http://www.hdpio.hr/savjetovanja/info99/okrugli-stol/papes/index.html) (<http://www.hdpio.hr/savjetovanja/info99/okrugli-stol/papes/index.html>)

Contact:

zlatko@mef.hr, davorin_bengez@yahoo.com, vkpapes@vip.hr
tel. 3769-079, 091-252-0324