Standardized multimedia elements in HTML5

Position paper for the W3C Video on the web workshop

Kevin Calhoun, Eric Carlson, Adele Peterson, Antti Koivisto

Apple Inc.

November 2007

1 Introduction

We believe that standard markup of multimedia – video and audio – as proposed in HTML5 [1] is important, and in this paper we explain some of the reasons that it is important, and explore some of the rationale and thinking behind what is currently proposed. We also will show, using the open-source WebKit project [6], examples of illustrating the advantages and opportunities that integrated support brings. This paper discusses material either proposed to the W3C [2] or in current W3C drafts, with additionally some open questions.

This paper does not discuss the nature of the audio/video coding or its container format. We agree that standardizing some format(s) is desirable for interoperability. However, the problems and issues raised by that question are not the subject of this paper, but are being handled, we believe, by the W3C staff.

We also contend that the standardized support of multimedia elements at the HTML5 level is valuable even if we cannot immediately settle on audio/video and container formats.

2 The state in HTML now

Audio and video are embedded in HTML today using either the <embed> or <object> elements. These have a number of problems.

First, they are not specific to multimedia, and so there is no general, uniform support for those characteristics shared by all multimedia elements. Thus, concepts as straightforward as the control of playback rate vary significantly between different plug-ins. This variation in capability extends across:

- the attributes available and their possible values;
- the user interface affordances presented (if any);
- varying document object model and scriptability;
- varying integration with styling and CSS [5];
- varying screen real-estate needs (e.g. for borders or controllers);
- varying provision for accessibility;

- poor and complex fallback handling, or the need for capability-probing and dynamically selected or generated HTML;
- varying support for composition; composition support varies on the browser/plug-in combination, not just on the plug-in.

These issues are exacerbated by the complication that a page author may be uncertain which plug-in will be used for 'public' MIME types, making the authoring of a page that uses a publicly specified (e.g. standard) multimedia type doubly difficult.

This level of complication results in authors either spending much effort to manage this complexity – effort that yields no obvious benefit to the page user – or avoiding it, and authoring pages with poor interoperability, behavior, accessibility, etc. Neither is desirable.

3 The proposal in HTML5

The basic proposal in HTML5 is that the HTML page be treated as an atemporal container of media elements. Those media elements are <video> and <audio>, where the latter has no visual aspect (this matches the semantics of video/ and audio/ top-level MIME types).

These multimedia elements have standardized attributes, behavior, DOM, styling etc.

They are similar to the image element, in that they can embed a variety of multimedia formats. Unlike image, they have explicit provision for fallback.

The treatment of multimedia is deliberately simple; there is provision for neither drawing (such as provided in SVG [4]) nor synchronization (such as in SVG and SMIL [3]). Instead, if these are needed, then suitable formats can be embedded using these elements, such as:

- SVG if drawing is needed;
- SMIL or SVG if synchronization and time behavior need to be specified.

It is important to note that the current proposal is precisely to provide a standard way to embed these document types in HTML5, with as little overlap as possible with these existing specifications, instead deferring to them if their capability is needed.

4 Advantages of the proposal

4.1 Standard markup

This includes the names of the elements, the names of their attributes, and the value-space that the attributes can take. Standardized markup at this level may seem basic, but simply harmonizing the needless small differences between media plug-ins greatly simplifies web authoring, and reduces the possibility of error.

The standardized markup extends to standard fallback behavior. Rather than the fallback affecting everything (as is the case with <object>), the common markup is expressed once, and only the actual source is subject to fallback.

4.2 Full integration with CSS and Media Queries

The use of media queries on the <source> elements allows for sophisticated fallback, handling cases not only of varying display or presentation capabilities, but also handling accessibility needs, by making it possible to mark particular sources as explicitly suitable (or unsuitable) for particular accessibility needs. In addition, accessibility would be enhanced by being able to style important aspects of the multimedia, such as its default playback rate.

Treating multimedia elements as first-class means that they can also get full CSS styling, including, for example, rollover behavior, opacity, and general styling (even for something as simple as display size or audio volume).

CSS also brings with it the ability to handle time-based transforms and animations, using the proposal we have made to CSS; their use uniformly across multimedia, images, and text, greatly enhances the web experience.

Finally, having drawing fully integrated into the browser means that not only CSS opacity, but also content-embedded opacity such as alpha coding, can be handled.

4.3 Uniform Document Object Model (DOM)

Providing a uniform DOM means that page authors can both interrogate and manipulate the multimedia content in a standard and uniform way. This again has accessibility ramifications.

4.4 Accessibility

The adoption of the accessibility guidelines for, and the accessibility of, time-based media, are in arguably the poorest condition for any web-accessible material. Much of this is caused by the fact that embedding even non-accessible content, and authoring a page to manage it, can be complex, as noted above; the complexity expands when accessibility is also considered.

Standardized markup and DOM means that accessibility at the ‘controller’ level can be enhanced, either by using the DOM to provide accessible controls of a particular kind, or indeed by special browser provision if desired.

Media queries, styling, and DOM also make accessibility at the content level more manageable, by selecting and then controlling or styling the content to provide the desired accessibility. This extends, for example, to controlling the default playback rate, which some users like to be slower than normal to handle issues of visual acuity or rate of comprehension.

4.5 Linking

The web is a web because it has cross-links. This may seem obvious, but making linking multimedia content is therefore important.

Given standardized browser behavior and DOM, linking *into* multimedia content should be much simpler (though, despite the work of MPEG for example, more work needs to be done in the content formats on fragment syntaxes for multimedia formats).

The provision of cue-ranges – time-spans which generate an entry and exit event as the playback time traverses them – greatly simplifies the linking *out of* multimedia content.

5 Open questions

5.1 Introduction

This section outlines some questions and issues *not* currently covered by the HTML5 draft.

5.2 Temporal Container

As noted above, the current design defers to embedded SMIL or SVG should temporal container semantics be needed, and explicitly treats HTML5 as an atemporal container. This works if the content that needs temporal container semantics is co-located on the page, and can be handled by a single embed of SMIL or SVG. However, if separated multimedia elements need synchronization, the use of cue ranges (the only current alternative) may not be sufficiently accurate. In addition, it may be desirable to synchronize animations or transformations with timed media.

5.3 Cue Ranges

Cue ranges today are functional: they are DOM manipulations. This means that they can neither be expressed in the markup directly (e.g. as values of one or more attributes) nor in the content itself. Given that some formats do have provision for similar concepts (e.g. AIFF ranges, QuickTime chapters), and that for some cue-ranges it may be more natural to express them in the format (e.g. chapters or other characteristics of the media), this may need examining.

5.4 Metadata

Many multimedia container formats handle static or time-parallel metadata, and there are also formats for expressing metadata outside the container format (e.g. in XML). It is probably desirable to expose this metadata in a standard way, so that, for example, browsers and pages can handle basic questions as the copyright status of multimedia.

However, there are serious questions here about security and the possibilities for cross-site scripting to expose material across security boundaries, and also difficult questions about the nature of a metadata-format-independent interface to metadata. One needs to be able to enquire what, for example, the title of the work is, without caring whether that is expressed in ID3, SMPTE KLV (key-length-value), MPEG-7, or any number of vendor-specific ways.

5.5 Embedding SMIL or SVG

As noted above, the current design defers to SMIL and SVG for some cases. However, we are not sure if an analysis has been done on whether all the advantages above apply equally when a video, for example, is embedded directly in HTML5 and when it is embedded in a SMIL or SVG document which in turn is embedded in HTML5.

5.6 Accessibility

There is much in here that makes better accessibility either automatically available (e.g. the standardized DOM), or easier to author and make available (e.g. selecting and styling content). Some aspects of accessibility may need more coverage, however. For example, some users like higher contrast on the video; this may be better handled through styling and/or DOM manipulation than through media-query selection of content authored with higher contrast.

In addition, there are probably questions of styling sub-parts of the multimedia content (e.g. rendering styles for subtitles or closed captions).

6 References

[1] Multimedia support in the HTML5 draft. <http://www.w3.org/html/wg/html5/> sections 3.14.7 through 3.14.10.

[2] Apple proposal on CSS and media elements <http://webkit.org/specs/Timed_Media_CSS.html>

[3] Synchronized Multimedia <http://www.w3.org/AudioVideo/>

[4] Scalable Vector Graphics (SVG) <http://www.w3.org/Graphics/SVG/>

[5] Cascading Style Sheets <http://www.w3.org/Style/CSS/>

[6] The WebKit Project <http://www.webkit.org/>