UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.,
AMAZON.COM SERVICES LLC,
AMAZON WEB SERVICES, INC., and
AUDIBLE, INC.,
Petitioners,

v.

AUDIO POD IP, LLC,
Patent Owner.

Case No. IPR2025-01003
U.S. Patent No. 9,729,907

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,729,907

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases: ***Page(s):***

Statutes and Rules:

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 9,729,907 ("the '907 patent")
1002	Declaration of Professor Ketan Mayer-Patel, Ph.D.
1003	NATIONAL INFORMATION STANDARDS ORGANIZATION, SPECIFICATIONS FOR THE DIGITAL TALKING BOOK (ANSI/NISO Z39.86-2002) (2002) ("DTB")
Exhibit Numbers 1004-1005 Not Used	
1006	Yoshimura et al., *Content Delivery Network Architecture for Mobile Streaming Service Enabled by SMIL Modification*, 86 IEICE TRANSACTIONS ON COMMC'N 1778 (2003) ("Yoshimura")
1007	Excerpts from DICK C.A. BULTERMAN & LLOYD RUTLEDGE, SMIL 2.0, INTERACTIVE MULTIMEDIA FOR WEB AND MOBILE DEVICES (2004) ("Bulterman")
Exhibit Number 1008 Not Used	
1009	U.S. Patent Publication No. 2003/0061305 ("Copley")
Exhibit Numbers 1010-1012 Not Used	
1013	U.S. Patent Publication No. 2003/0033147 ("McCartney")
1014	U.S. Patent Publication No. 2003/0081336 ("Woodill")
1015	U.S. Patent No. 5,857,203 ("Kauffman")
Exhibit Numbers 1016-1026 Not Used	
1027	U.S. Patent Publication No. 2003/0013073 ("Duncan")
Exhibit Number 1028 Not Used	

Exhibit No.	Description
1029	Robert Steele et al., *Accessing SMIL-based Dynamically Adaptable Multimedia Presentations from Mobile Devices*, in INT'L CONF. ON INFO. TECH.: CODING AND COMPUTING (2004) ("Steele")
1030	U.S. Patent No. 6,260,011 ("Heckerman")
Exhibit Number 1031 Not Used	
1032	U.S. Patent. No. 6,108,703 ("Leighton")
1033	U.S. Patent No. 5,922,045 ("Hanson")
Exhibit Numbers 1034-1044 Not Used	
1045	Chun-Chuan Yang et al., *Design and Implementation of the Just-in-Time Retrieving Policy for Schedule-Based Distributed Multimedia Presentations*, 71 J. SYS. & SOFTWARE 49 (2004) ("Yang")
1046	Tredinnick, *Implementing the SMIL Specification*, PROC. OF THE LINUX SYMP. (2003) ("Tredinnick")
Exhibit Numbers 1047-1048 Not Used	
1049	UNIFORM RESOURCE IDENTIFIERS (URI): GENERIC SYNTAX, RFC 2396 (1998)
Exhibit Number 1050 Not Used	
1051	Excerpts from World Wide Web Consortium, *Synchronized Multimedia Integration Language (SMIL 2.0) Specification*, W3C Recommendation (Aug. 7, 2001), https://www.w3.org/TR/2001/REC-smil20-20010807/smil20.html. (the "SMIL Standard")
Exhibit Numbers 1052-1056 Not Used	
1057	U.S. Patent No. 5,798,841 ("Takahashi")

Exhibit No.	Description
1058	U.S. Patent No. 5,021,989 ("Fujisawa")
1059	PCT Patent Publication No. WO2001/01373 ("Hendricks")
Exhibit Numbers 1060-1062 Not Used	
1063	P. Delacourt & C.J. Wellekens, *DISTBIC: A speaker-based segmentation for audio data indexing*, 32 SPEECH COMM'N 111 (2000) ("Delacourt")
Exhibit Numbers 1064-1068 Not Used	
1069	PCT Patent Publication No. WO2002/080524 ("Dimitrova")
Exhibit Numbers 1070-1071 Not Used	
1072	Excerpt from File History for European Patent Publication No. EP1961154
Exhibit Number 1073 Not Used	
1074	Final Office Action dated July 28, 2010, U.S. Patent App. No. 12/096933
1075	Barry Arons, *SpeechSkimmer: A System for Interactively Skimming Recorded Speech*, 4 ACM TRANSACTIONS ON COMPUTER-HUMAN INTERACTION 3 (1997) ("Arons")
Exhibit Numbers 1076-1085 Not Used	
1086	PCT Patent Publication No. WO2001/024474 ("Shteyn")
Exhibit Numbers 1087-1094 Not Used	
1095	Excerpts from the File History for U.S. Patent No. 9,729,907
1096	CV of Professor Ketan Mayer-Patel, Ph.D.
1097	Declaration of Sylvia D. Hall-Ellis, Ph.D.
Exhibit Numbers 1098-1100 Not Used	

Exhibit No.	Description
1101	John E. Simpson, *From XML to SMIL*, XML.COM (May 26, 2004), https://web.archive.org/web/20040606014049/https:/www.xml.com/pub/a/2004/05/26/qa.html (June 6, 2004 snapshot via the Wayback Machine) ("Simpson")
1102	Ingo Elsen et al., *Streaming Technology in 3G Mobile Communication Systems*, 34 COMPUTER, no. 9, at 46-52 (2001) ("Elsen")
1103	WIKIBOOKS, *XML – Managing Data Exchange/SMIL*, https://en.wikibooks.org/w/index.php?title=XML_-_Managing_Data_Exchange/SMIL&oldid=283559 (November 15, 2005 snapshot via Wikibooks) ("SMIL Wiki")
1104	Lloyd Rutledge, *SMIL 2.0: XML for Web Multimedia*, IEEE INTERNET COMPUTING, Sept.-Oct. 2001, at 78-84 ("Rutledge")
1105	Dongkyoo Shin et al., *Design and Implementation of the SMIL (Synchronized Multimedia Integration Language) Player*, 48 IEEE TRANSACTIONS ON CONSUMER ELECS. 575 (2002) ("Shin")
1106	Herng-Yow Chen et al., *Design of a Web-Based Synchronized Multimedia Lecture System for Distance Education*, 2 PROC. IEEE INT'L CONF. ON MULTIMEDIA COMPUTING AND SYS. 887 (1999) ("Chen")
1107	U.S. Patent Publication No. 2007/0124331 ("Griffin")

Petitioners Amazon.com, Inc., Amazon.com Services LLC, Amazon Web Services, Inc., and Audible, Inc. ("Petitioners" or "Amazon") respectfully request *inter partes* review of claims 1-21 of U.S. Patent No. 9,729,907 ("the '907 patent"), which Audio Pod IP, LLC ("Patent Owner" or "PO") purportedly owns.

I. INTRODUCTION

The challenged claims relate to creating a file for synchronizing digital media streams (e.g., audio and text) from the same originating work. Such synchronization files had been in use for nearly a decade prior to the '907 patent's earliest possible priority date of December 13, 2005. By 2005, synchronization files were so common there were multiple industry standards implementing them. (EX-1002 ¶32.) For example, in the late 1990s, a standard was developed for a "Synchronized Multimedia Integration Language (SMIL)," which used XML files to synchronize text, audio, and images. (EX-1046, 424-25.) In 2002, the SMIL specification was used by the National Information Standards Organization in a standard for digital talking books ("DTBs") that provided synchronized rendering of digital versions of text, audio, and images from print books for the visually impaired. (EX-1003, vii, 6-7.)

Other synchronization files for "talking books"—now commonly referred to as "audiobooks"—were also well known before 2005. (EX-1002 ¶33.) For example, in 2003, McCartney disclosed a method "for coordinating independently produced text and audio clip data" to produce "a properly coordinated and constructed

digital talking book" using "synchronization files." (EX-1013, Abstract.)

In view of the wealth of prior art describing synchronization files to synchronize digital media streams, the '907 patent claims should never have issued.

II. THE '907 PATENT

A. Overview

The '907 patent describes creating audiobooks having synchronized media streams. As shown in Figure 24a below, a "presentation manager" combines media streams (e.g., audio, text, and illustrations) using synchronization information from a "Virtual Media Descriptor" to generate an audiobook:

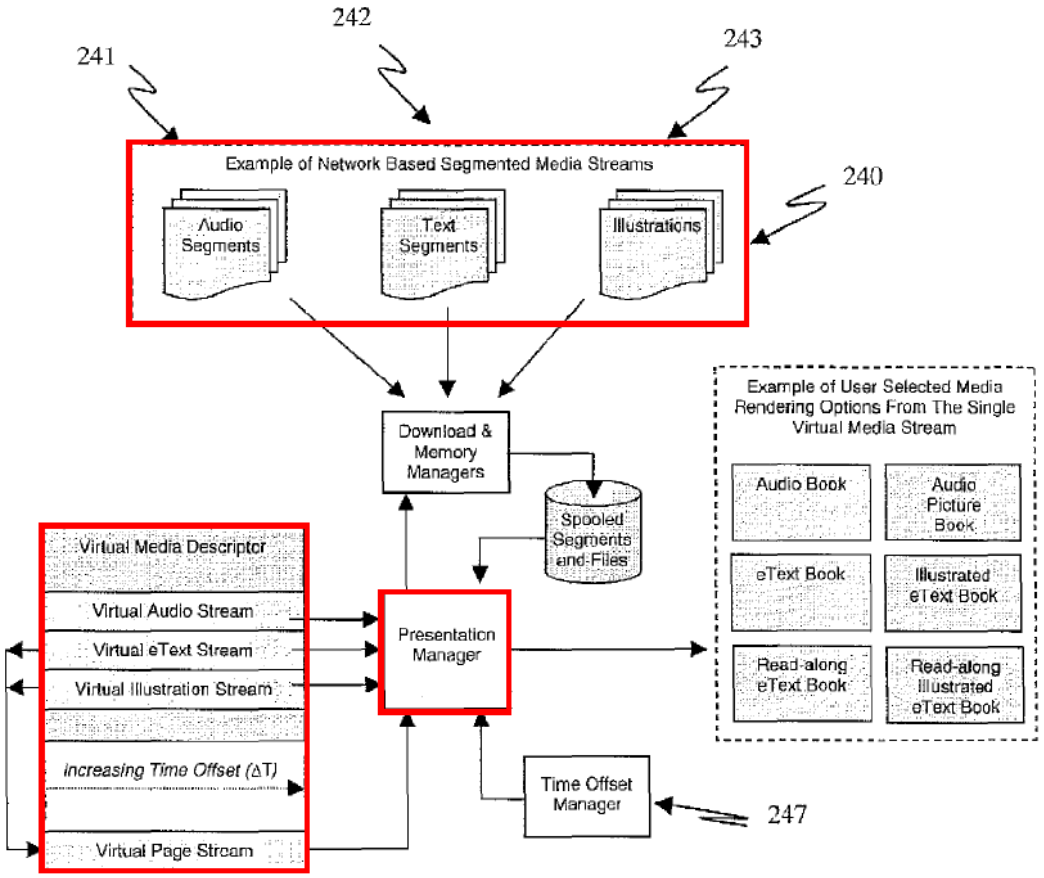


Fig. 24a

(EX-1001, Fig. 24a[1].)

The Virtual Media Descriptor includes a "Virtual Audio Stream" and a "Virtual eText Stream." (*Id.*, 20:23-39, 33:54-57.) The Virtual Audio Stream includes time information identifying points (e.g., the start of a chapter) in the audio stream. (*Id.*, 20:48-55.) This information is used to assign time information to the corresponding points (e.g., the start of the same chapter) in the text stream. (*Id.*, 27:9-18, 33:57-65.) The time information for the text content is stored in the Virtual eText Stream. (*Id.*, 26:60-27:7, 33:57-60.) Thus, the Virtual Audio Stream and Virtual eText Stream are "linked via a common timeline," which facilitates synchronization and navigation of text and audio streams "without modifying the media streams with identifiers." (*Id.*, 22:41-67, 33:39-42; EX-1002 ¶¶34-35.)

B. Prosecution

The Examiner allowed the claims without issuing a single Office Action. (EX-1095, 27-28.) None of the references relied on herein were submitted to, or considered by, the Examiner.

C. Priority

The patent's earliest possible priority date is December 13, 2005. (EX-1001, 1-2.) Petitioners do not concede that the claims are entitled to that priority date.

[1] Images herein may be colored and/or annotated for clarity.

III. LEVEL OF ORDINARY SKILL IN THE ART

A POSITA is "a person of ordinary creativity, not an automaton." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 421 (2007). Here, a POSITA would have had at least a bachelor's degree in electrical engineering, computer engineering, or computer science, and at least three years of industry or academic experience in the design, development, and/or implementation of content rendering and/or distribution systems. (EX-1002 ¶¶27-31); *see In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995). Work experience could substitute for formal education and additional formal education could substitute for work experience. (EX-1002 ¶29.)

IV. CLAIM CONSTRUCTION

No claim terms require construction to resolve the obviousness challenges here. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999). For purposes of this proceeding only, Petitioners assume the claims are not invalid under §112.

V. STATEMENT OF PRECISE RELIEF REQUESTED

A. Grounds

The Board should cancel the claims as obvious under 35 U.S.C. §103 on the following Grounds:

Ground	Challenged Claims	References
1A	1-14, 17-19, 21	DTB
1B	2, 15, 18-19	DTB and Yoshimura
1C	7	DTB and Duncan
1D	8	DTB and Heckerman
1E	14, 20	DTB and Bulterman
1F	15	DTB and Yang
1G	16	DTB, Yang, and Copley
2A	1-3, 8-12, 14, 17-19, 21	McCartney
2B	2, 18	McCartney and Kauffman
2C	4-6, 13	McCartney and DTB
2D	7	McCartney and Duncan
2E	8	McCartney and Woodill
2F	14, 20	McCartney and Bulterman
2G	15	McCartney and Shteyn
2H	16	McCartney, Shteyn, and Copley

Additional support is included in the Declaration of Professor Ketan Mayer-Patel, Ph.D. (EX-1002.)

B. Status of References as Prior Art

Each of the following references is prior art under pre-AIA §102(b) because it published more than one year before the '907 patent's earliest possible priority date of December 13, 2005: DTB (2002); Bulterman (May 2004); Shteyn (2001);

McCartney (2003); Kauffman (1999); Woodill (2003); Duncan (2003); Yoshimura (2003); Heckerman (2001); Copley (2003); and Yang (February 2004). (*See* EX-1097.)

These references are analogous art because each is from the same field of endeavor as the '907 patent, e.g., content distribution and/or rendering. (EX-1002 ¶22.) They are also pertinent to a particular problem the inventor was focused on, e.g., efficient and effective distribution and/or rendering of content. (*Id.*)

VI. GROUND 1A: CLAIMS 1-14, 17-19, AND 21 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB.

DTB describes methods by which the text of a book can be "synchronized with the audio presentation" using a SMIL file. (EX-1003, vii, 21.)

A. Claim 1

1. Preamble

The preamble recites "[a] method comprising." DTB discloses a method of creating digital talking books using SMIL files. (EX-1003, Abstract, 6; EX-1002 ¶38.)

2. 1[a][i]: Descriptor File

Element 1[a][i] recites "creating a descriptor file for synchronizing a plurality of digital media streams."

DTB discloses creating a descriptor file, such as a "Synchronized Multimedia Integration Language [SMIL]" file, for synchronizing digital audio, text, and image

streams to provide a "digital talking book." (EX-1003, 3; *id.*, 21, 61 ("[t]ext, audio, and [] images are synchronized in the SMIL files"), 23-43, vii (DTB enables synchronization of digital media streams), 4, 6, 7-9, 12; EX-1002 ¶40.)

Thus, DTB discloses creating a descriptor file (e.g., SMIL file) for synchronizing a plurality of digital media streams (e.g., text, audio, and image streams). (EX-1002 ¶¶39-41.)

3. 1[a][ii]-1[a][iii]: Digital Media Streams

Elements 1[a][ii]-1[a][iii] recite that "the plurality of digital media streams each contain digital media content corresponding to a same originating work" and "the plurality of digital media streams includes a first digital media stream containing a digital audio narration of the originating work and one or more other digital media streams." "An Originating Work is the complete text of a printed document, such as an article, speech, book or other written work[.]" (EX-1001, 19:40-46, 19:56-60.)

DTB discloses digital media streams containing the audio narration of an originating work (e.g., book). For example, DTB's audio files "can include human or synthetic speech recordings" of the complete originating work. (EX-1003, 6-9, 12 ("narrator" attribute includes the "[n]ame of the person whose recorded voice is embodied in the DTB"), 66 (SMIL file "permit[s] concurrent display of the text being narrated"), 60-61, vii, Abstract, vii, 4.)

DTB also discloses at least one other digital media stream, such as the complete "text" and/or "images" of the originating work (e.g., book). (EX-1003, 6 (DTB includes "a combination of audio, text, and images"), 6 ("In addition to text and audio, DTBs can include images"), 60-61 (DTB may include "full text of the document"); EX-1002 ¶45.)

Thus, DTB discloses that the plurality of digital media streams each contain digital media content (e.g., text, images, or audio) corresponding to a same originating work (e.g., print material including reference works, text books, magazines, etc.) and the plurality of digital media streams includes a first digital media stream containing a digital audio narration of the originating work and one or more other digital media streams (e.g., text and/or image). (EX-1002 ¶¶42-46.)

4. 1[a][iv]: External Descriptor File

Element 1[a][iv] recites "wherein the descriptor file is external to the first digital media stream."

DTB's SMIL file (e.g., descriptor file) (highlighted below) is separate from (e.g., external to) to the **audio stream** (e.g., the first digital media stream) (blue) with different extensions (e.g., ".smil" vs. ".mp3"):

```
<manifest>
     <item id="opf" href="rs.opf" media-type="text/xml" />
     <item id="text" href="rs.xml" media-type="text/xml" />
     <item id="text_style" href="dtbbase.css" media-type="text/css2" />
     <item id="ncx" href="rs.ncx" media-type="text/xml" />
     <item id="ncx_style" href="ncx16.css" media-type="text/css2" />
     <item id="SMIL" href="rs.smil" media-type="application/smil" />
     <item id="foreword" href="rs_fwdx.mp3" media-type="audio/mp3" />
     <item id="standards" href="rs_stdx.mp3" media-type="audio/mp3" />
     <item id="appendices" href="rs_app.mp3" media-type="audio/mp3" />
     <item id="index" href="rs_index.mp3" media-type="audio/mp3" />
     <item id="fig_01" href="fig1.png" media-type="image/png" />
     <item id="resource" href="rs.res" media-type="text/xml" />
     <item id="resource_audio" href="res.mp3" media-type="audio/mp3" />
</manifest>
```

(EX-1003, 14-15; *id.*, 27 (audio streams are "point[ed] to" by SMIL file).) Thus, DTB discloses this limitation. (EX-1002 ¶¶47-50.)

5. 1[b]: Location Information

Element 1[b] recites "storing location information for the plurality of digital media streams in the descriptor file."

DTB's SMIL files store location information for the plurality of media streams as source information including a uniform resource identifier ("URI"):

- <audio>
  Description: Points to segment of audio content to be rendered.
  Declaration: `<!ELEMENT audio EMPTY >`
  Syntax: `<audio...attributes... />`
  Attributes:
  - id (ID, IMPLIED): Optional identifier.
  - src (CDATA, REQUIRED): URI of audio file containing clip to be rendered.

- <text>
  Description: Points to segment of textual content to be rendered.
  Declaration: `<!ELEMENT text EMPTY >`
  Syntax: `<text ...attributes... />`
  Attributes:
  - id (ID, IMPLIED): Optional identifier.
  - src (CDATA, REQUIRED): URI of fragment of textual content file to be rendered.

- <img>
  Description: Points to image to be rendered.
  Declaration: `<!ELEMENT img EMPTY >`
  Syntax: `<image...attributes... />`
  Attributes:
  - id (ID, IMPLIED): Optional identifier.
  - src (CDATA, REQUIRED): URI of image file to be rendered.

(EX-1003, 25-27; *see also id.*, 98, 19, 3, 32; EX-1002 ¶52.)

DTB's URI is a locator because it "[p]oints to," for example, a "segment of audio content to be rendered." (EX-1003, 25-27.) Indeed, DTB discloses that the "src" attribute "specifies by URI *the location* of" the file. (*Id.*, 98 (emphasis added); EX-1002 ¶53.) POSITAs understood that src attributes in SMIL files are "used for locating and fetching the associated media." (EX-1051, 108; EX-1007, 90-91, 86; *see* EX-1049, 2; EX-1002 ¶53.)

Accordingly, DTB discloses or renders obvious storing location information (e.g., URI in src attribute) for the plurality of digital media streams (e.g., audio, text, image streams) in the descriptor file (e.g., SMIL file). (EX-1002 ¶¶51-54.)

Alternatively, if "location information" includes a location within the file (e.g., play location), DTB discloses this limitation because its SMIL file includes (1) clipBegin and clipEnd attributes specifying "time-based positions" in the audio and

(2) URIs identifying the corresponding position within the textual content. (EX-1003, 21-24, 3, 6-7, 27; EX-1002 ¶55.)

6. 1[c][i]: Time Offsets

Element 1[c][i] recites "identifying a plurality of time offsets in a timeline of the digital audio narration of the first digital media stream." "Time offset" "refers to the period of time, or difference in time, between two points in time[.]" (EX-1001, 21:60-62.)

DTB's audio "files are indexed by time-based positions[.]" (EX-1003, 21.) The time-based positions are represented as **clipBegin** and **clipEnd** attributes (pink), which specify the beginning and end of a segment of an **audio file** (blue):

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
   clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>
```

(*Id.*, 32; *see also id.*, 31-34.) The clipBegin and clipEnd attributes specify "a time offset from the start of the audio file":

- <audio>
  Description: Points to segment of audio content to be rendered.
  Declaration: `<!ELEMENT audio EMPTY >`
  Syntax: `<audio...attributes... />`
  Attributes:
  - **id (ID, IMPLIED):** Optional identifier.
  - **src (CDATA, REQUIRED):** URI of audio file containing clip to be rendered.
  - **type (CDATA, IMPLIED):** Type of media file.
  - **clipBegin (CDATA, IMPLIED):** Specifies the beginning of a segment of a continuous audio file as a time offset from the start of the audio file. The value syntax is defined by the SMIL 2.0 Timing and Synchronization Module [SMIL] See Section 7.7, "Clock Values."
  - **clipEnd (CDATA, IMPLIED):** Specifies the end of a segment of a continuous audio file as a time offset from the start of the audio file. It uses the same attribute value syntax as `clipBegin`.
  - **region (CDATA, IMPLIED):** Specifies the `region` (defined in `layout` in document `head`) in which the audio object will be presented. References the id of the appropriate `region`.

  Valid inside: `body, par, seq`

(*Id.*, 27; EX-1002 ¶59.) The offsets are "identified" at least when they are indexed or otherwise determined by the DTB producer. (EX-1003, 21; EX-1002 ¶59.)

These offsets are also "in a timeline of the digital audio narration of the first digital media stream." (EX-1002 ¶60.) When DTB's audio stream is a single file, the offsets represent the time from the beginning of the stream to the selected point. (EX-1003, 37 (an audio file ordinarily contains the entire audio content of DTB); EX-1002 ¶60.) When the stream comprises multiple audio files, the time offsets within that file are time offsets in *a* timeline (e.g., an audio file) of the first digital media stream. (*Id.*)

Accordingly, DTB discloses or renders obvious identifying (e.g., determining and/or indexing) a plurality of time offsets (e.g., clipBegin and/or clipEnd times) in a timeline of the digital audio narration of the first digital media stream (e.g., timeline of the audio file). (EX-1002 ¶¶57-61.)

7. 1[c][ii]: Content Points in Audio Narration

Element 1[c][ii] recites "wherein the plurality of time offsets correspond to a plurality of content points in the digital audio narration." The '907 patent refers to the beginning of a chapter, paragraph, or sentence as examples of "points" in the digital media stream. (EX-1001, 21:42-45.)

SMIL files group media content into "parallels" or <par> elements. (EX-1003, 26, 27 (parallel is a "[p]arallel time grouping in which multiple elements (e.g., text, audio, and image) play back simultaneously"); EX-1002 ¶64.) Each parallel comprises portions of media streams that comprise the same "body of material (e.g., the same paragraph)" or heading. (EX-1003, 29-30, 26, 32.) These bodies of material (e.g., paragraphs or headings) are content points. (EX-1002 ¶64.) For example, the SMIL file below shows a <par> for two **content points** (green): a heading ("**h1**") and a paragraph ("**para1**"):

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
   clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>
```

(EX-1003, 32.) Thus, DTB discloses that the plurality of time offsets (e.g., clipBegin and clipEnd) correspond to a plurality of content points (e.g., headings and/or paragraphs) in the digital audio narration. (EX-1002 ¶¶62-65.)

8. 1[d]: Correlating Time Offsets and Content Points

Element 1[d] recites "storing the plurality of time offsets and the plurality of content points in the descriptor file in a manner indicating a correlation between the plurality of time offsets and the plurality of content points."

DTB's plurality of time offsets (e.g., clipBegin and clipEnd) (pink) are contained in content tags (e.g., <audio> tags) within the parallel corresponding to the content point (e.g., heading, paragraph number, etc.) (green):

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
   clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>
```

(EX-1003, 32; *see also id.*, 26; EX-1002 ¶67.)

Thus, DTB discloses storing the plurality of time offsets (e.g., clipBegin and/or clipEnd) and the plurality of content points (e.g., h1, para1, etc.) in the descriptor file in a manner indicating a correlation between them. (EX-1002 ¶¶66-68.)

9. 1[e]: Synchronization Points

Element 1[e] recites "identifying synchronization points in the digital media content of the one or more other digital media streams." The patent explains that "a time offset can be calculated for each word" such that synchronization can occur for

"words, phrases, lines, sentences, paragraphs, or illustrations[.]" (EX-1001, 31:4-17.)

DTB "associat[es] time-points in the audio file with the corresponding positions within the textual content" or image stream using the SMIL file's parallels. (EX-1003, 21, 26.) Synchronization can occur "down to paragraph, sentence, or even word level." (*Id.*) DTB expressly refers to these points as "synchronization point[s]." (*Id.*, 21.)

For example, the SMIL file below shows several synchronization points, highlighted in teal. (*Id.*, 32; EX-1002 ¶72.) The "h1" parallel synchronizes heading 1 ("h1_1") of the text file (purple) with "00:00:01.62"-"00:00:02.53" of the audio file. (*Id.*) The "para1" parallel synchronizes paragraph 1 ("para_1") of the text file (purple) with "00:00:03.51"-"00:01:45.36" of the audio file. (*Id.*) And the "img1" parallel synchronizes the image ("fig1.png") (orange) and the associated caption ("caption_1") of the text file (purple) with "00:01:45.98"-"00:01:52.66" of the audio file. (*Id.*)

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
   clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>

<par id="img1">
   <img region="image" src="fig1.png" />
<seq id="icap1">
      <par id="cap1">
         <text region="caption" src="rs.xml#caption_1" />
         <audio src="rs_fwdx.mp3"
         clipBegin="00:01:45.98" clipEnd="00:01:52.66" />
      </par>
      <par id="pnote1" customTest="prodnote">
         <text region="text" src="rs.xml#prodnote_1" />
         <audio src="rs_fwdx.mp3" clipBegin="00:01:53.08"
         clipEnd="00:02:55.34" />
      </par>
   </seq>
</par>
```

(*Id.*)

The synchronization points are determined by a DTB producer. (*Id.*, 21.) Accordingly, DTB discloses or renders obvious identifying synchronization points (e.g., h1_1, para_1, fig1.png, and/or caption_1) in the digital media content of the one or more other digital media streams (e.g., text and/or image stream). (EX-1002 ¶¶69-74.)

10. 1[f]: Synchronization Time Offsets

Element 1[f] recites "selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets."

When a text or image segment is synchronized with the audio stream, the time offsets identified in the clipBegin and clipEnd attributes contained in the <audio> tag within a parallel are "synchronization time offsets." (EX-1003, 61 (when present, text is "synchronized with the corresponding audio" in the SMIL files); EX-1002 ¶76.) If, however, there is no corresponding text segment or image in the parallel, then the time offsets are not "synchronization time offsets" because they do not synchronize corresponding media streams. (*Id.*; *see also* EX-1003, 60-61.) Moreover, because "[t]he DTB producer determines granularity of the synchronization events," the synchronization time offsets are "selected." (*Id.,* 21; EX-1002 ¶76.) Thus, DTB discloses or renders obvious selecting synchronization time offsets (e.g., clipBegin and/or clipEnd attributes associated with text segments or images) that correspond to the synchronization points (e.g., h1_1, para_1, fig1.png, caption_1) from the plurality of time offsets (e.g., all available clipBegin and clipEnd attributes). (EX-1002 ¶¶75-77.)

11. 1[g][i]: Correlating Synchronization Time Offsets and Synchronization Points

Element 1[g][i] recites "storing the synchronization time offsets and the synchronization points in the descriptor file in a manner indicating a correlation between the synchronization time offsets and the synchronization points."

As illustrated below, DTB's synchronization time offsets (e.g., clipBegin and clipEnd) (red) and their corresponding synchronization points (e.g., synchronized

text and/or images) (teal) are stored in the descriptor file (SMIL file) within the same parallel, indicating a correlation between them:

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
clipEnd="00:01:45.36" />
</par>

<par id="img1">
   <img region="image" src="fig1.png" />
   <seq id="icap1">
      <par id="cap1">
         <text region="caption" src="rs.xml#caption_1" />
        <audio src="rs_fwdx.mp3"
         clipBegin="00:01:45.98" clipEnd="00:01:52.66" />
      </par>

      <par id="pnote1" customTest="prodnote">
         <text region="text" src="rs.xml#prodnote_1" />
         <audio src="rs_fwdx.mp3" clipBegin="00:01:53.08"
         clipEnd="00:02:55.34" />
      </par>
   </seq>
</par>
```

(EX-1003, 32; EX-1002 ¶79.) Thus, DTB discloses this limitation. (EX-1002 ¶¶78-80.)

12. 1[g][ii]: Synchronized Rendering

Element 1[g][ii] recites "such that the descriptor file allows a synchronized rendering of the plurality of digital media streams on a client device."

DTB's SMIL file enables "synchronized presentation" of audio, text, and images on DTB players. (EX-1003, 21; *see also id.*, 3, 6-7, 61; *supra* §VI.A.2; EX-

1002 ¶¶81-83.)

Thus, DTB renders claim 1 obvious. (EX-1002 ¶¶38-84.)

B. Claim 2

Claim 2 depends from claim 1 and further recites "storing the descriptor file on a server that is accessible to the client device."

DTB discloses playback of multimedia presentations "over the Internet," which a POSITA would have understood involves retrieving content not originally resident on the client device. (EX-1003, 21; EX-1002 ¶86.) It would have been obvious to a POSITA to store the descriptor file on a server for multiple reasons.

First, storing a SMIL file on a client accessible server was widely known. (*E.g.*, EX-1086, 3:33-34; EX-1032, 4:57-67; EX-1002 ¶87.)

Second, a POSITA would have understood the benefits of storing the SMIL file on a client accessible server, including the ability to update SMIL files to, for example, indicate the best cache server location, improving streaming experience. (EX-1006, 1779, 1781.) It would also allow for centralized management and version and access control of SMIL files. (EX-1002 ¶88.) A POSITA would have reasonably expected success in storing files on client-accessible servers because doing so was well known, disclosed in many prior art references, and routine to a POSITA. (EX-1002 ¶89 (citing EX-1086, 3:19-4:2; EX-1102, 50; EX-1107 ¶[0013]).)

C. Claim 3

Claim 3 depends from claim 1 and further recites "simultaneously rendering two or more of the plurality of digital media streams on the client device by using the descriptor file to synchronize the two or more of the plurality of digital media streams."

DTB discloses this limitation. (*See, e.g.*, EX-1003, 7; *supra* §VI.A.12; EX-1002 ¶¶92-93.)

D. Claim 4

Claim 4 depends from claim 3. Element 4[a] recites "tracking a current position in at least one of the two or more of the plurality of digital media streams as the two or more of the plurality of digital media streams are rendered" and element 4[b] recites "creating a bookmark by setting the current position as a bookmarked position wherein the bookmark includes the bookmarked position and identifies the descriptor file."

DTB players "use the synchronization information" in the SMIL file to "track, during audio playback, the corresponding position in the textual content file[.]" (EX-1003, 7.) This information can then be used in a "bookmark file." (*Id.*, 45-50 (bookmark can be set at "any point in a DTB, whether based on the audio file or the textual content file").) The bookmark file includes a <bookmark> element that represents the "[p]oint in [a] document marked by user for direct access in the future"

and/or a <lastmark> element that identifies the "[l]ocation where user most recently ceased reading[.]" (*Id.*, 47.)

As illustrated below, the <bookmark> and <lastmark> elements of DTB's bookmark file each include (1) a URI that identifies the SMIL file (e.g., descriptor file) containing the position (red) and (2) a time offset to the exact point in the <bookmark> or <lastmark> element (e.g., bookmarked position) (blue). (EX-1003, 47-48; EX-1002 ¶96.)

```
<title>
  <text>Gone with the Wind</text>
  <audio src="gwtw_title.mp3" />
</title>
<uid>us-rfbd-JT065</uid>

<lastmark>
  <ncxRef>gwtw.ncx#lvl1_5</ncxRef>
  <uri>gwtw_ch5.smil#para023</uri>
  <timeOffset>173</timeOffset>
</lastmark>

<bookmark>
  <ncxRef>gwtw.ncx#lvl1_1</ncxRef>
  <uri>gwtw_ch1.smil#para008</uri>
  <timeOffset>22</timeOffset>
</bookmark>
```

(EX-1003, 49; EX-1002 ¶96.) The position may also be identified by a character offset. (EX-1003, 47, 50; EX-1002 ¶96.)

Thus, DTB discloses or renders obvious tracking a current position in at least one of the two or more of the plurality of digital media streams (e.g., audio or text streams) as the two or more of the plurality of digital media streams are rendered and creating a bookmark (e.g., <bookmark> or <lastmark> element) by setting the current position as a bookmarked position (e.g., playback position when marked) wherein the bookmark includes the bookmarked position (e.g., time or character off-set) and identifies the descriptor file (e.g., SMIL file). (EX-1002 ¶¶94-98.)

E. Claim 5

Claim 5 depends from claim 4 and further recites "rendering one or more of the plurality of digital media streams starting from the bookmarked position on the client device or on a second other client device."

DTB players can resume play at either the <bookmark> or <lastmark> position on the playback device. (*Supra* §VI.D; EX-1003, 47-48; *id.*, 45-50.) Moreover, because bookmark files are "exportable," it would have been obvious to a POSITA to use them to resume playback from the bookmarked position(s) on a second other client device. (EX-1003, 7; EX-1002 ¶100 (citing EX-1107 ¶[0005]).)

F. Claim 6

Claim 6 depends from claim 5 and further recites "wherein the one or more of the plurality of digital media streams were not used to create the bookmark."

DTB bookmarks are "based on the audio file *or* the textual content file." (EX-1003, 45 (emphasis added), 49-50 (disclosing bookmark using audio stream and lastmark using text stream).) Thus, DTB discloses that one digital media stream (e.g., text) is not used to create the bookmark. (*Id.*) DTB also discloses an image stream, but it is not used to create the bookmark. (*Id.*; EX-1002 ¶¶102-04.)

G. Claim 7

Claim 7 depends from claim 3.

1. 7[a]: Visually Rendered Content

Element 7[a] recites that "at least one of the two or more of the plurality of digital media streams contains visually rendered digital media content."

DTB discloses this limitation. (EX-1003, 6, vii; *supra* §VI.A.3; EX-1002 ¶105.)

2. 7[b]: Automated Page Turns

Element 7[b] recites "providing automated page turns as the two or more of the plurality of digital media streams are simultaneously rendered."

DTB discloses that "page numbers" can be "presented … during normal play-back of a DTB." (EX-1003, 25; *see also id.*, 95 (page number "inserted at the point within the file immediately preceding the first item of content on a new page").) Thus, DTB discloses or renders obvious providing automated page turns as the two

or more of the plurality of digital media streams (e.g., text and audio streams) are simultaneously rendered (e.g., during playback). (EX-1002 ¶¶106-09.)

H. Claim 8

Claim 8 depends from claim 1 and further recites "wherein at least some of the synchronization time offsets are selected in dependence on natural language gaps, inter-word gaps, punctuation marks, or parts of speech."

DTB discloses selecting synchronization time offsets down to the "paragraph, sentence, or even word level." (*Supra* §§VI.A.9-VI.A.11.) A POSITA would therefore have understood DTB to disclose selecting synchronization time offsets in dependence on natural language gaps (e.g., sentence and paragraph level), punctuation marks (sentence level) or inter-word gaps (word level). (EX-1002 ¶¶110-12.)

I. Claim 9

Claim 9 depends from claim 1 and further recites "wherein the first digital media stream includes a plurality of content segments, and wherein the plurality of content segments are defined by the plurality of time offsets in the descriptor file."

DTB's audio file (e.g., "rs_fwdx.mp3") may be divided into audio "segment[s]" or "clips" that are represented as <audio> elements and defined by clipBegin and clipEnd offsets:

```
<par id="h1">
   <text region="text" src="rs.xml#h1_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:01.62"
   clipEnd="00:00:02.53" />
</par>

<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>
```

(EX-1003, 31-34, 27; *see also id.*, 33-34, 6 (narration embodied in multiple audio files), 15.)

Thus, DTB discloses or renders obvious that the first digital media stream (e.g., audio stream) includes a plurality of content segments (e.g., different audio clips and/or files), and wherein the plurality of content segments are defined by the plurality of time offsets (e.g., clipBegin and clipEnd) in the descriptor file (e.g., SMIL file). (EX-1002 ¶¶113-16.)

J. Claim 10

Claim 10 depends from claim 9 and further recites "wherein the plurality of time offsets includes start times of the content segments, end times of the content segments, durations of the content segments, or a combination thereof."

The clipBegin and clipEnd times reflect the start and end time of the content segment identified in each parallel. (EX-1003, 27, 31-34; EX-1002 ¶¶117-19.)

K. Claim 11

Claim 11 depends from claim 1.

1. 11[a]: Selecting Streams

Element 11[a] recites "selecting two or more of the plurality of digital media streams."

A DTB user selects media streams by selecting a book and/or jumping to a particular section (e.g., a chapter, heading, section, page, figure, etc.) using bookmarks or lastmarks. (EX-1003, 55, 41; EX-1002 ¶121.) In response, the player selects two or more of the digital media streams (e.g., audio and text) for playback. (EX-1002 ¶121; EX-1003, 61-62.) Thus, DTB discloses or renders obvious this limitation. (EX-1002 ¶¶120-21.)

2. 11[b]: Determining a First Position

Element 11[b] recites "determining a first position in the digital media content of the two or more of the plurality of digital media streams in dependence on the synchronization time offsets in the descriptor file."

DTB discloses determining a position using the synchronization time offsets both (1) during playback and (2) to access a bookmark/lastmark position. (EX-1002 ¶122.)

First, DTB players use the synchronization information, including offsets in the audio file, to track the corresponding position in the text and image files during playback. (EX-1003, 7; *supra* §VI.D; EX-1002 ¶123.)

Second, the bookmark or lastmark points to the relevant <par> or <seq> in the SMIL file and provides a time offset relative to the clipBegin of that portion. (EX-1003, 46-50.) The position of the bookmark or lastmark is therefore determined in the audio stream in dependence on the synchronization time offsets (e.g., clipBegin) in the descriptor file (SMIL file). (EX-1002 ¶124.) Because the <par> elements are "[p]arallel time grouping[s] in which multiple elements (e.g., text, audio, and image) play back simultaneously[,]" the determination of a location within the audio file using a <par> element of a SMIL file also determines the position of any corresponding text or image streams. (EX-1003, 6-7, 21-27, 60-61; EX-1002 ¶124.)

Thus, DTB discloses or renders obvious this limitation. (*Id.* ¶¶122-25.)

3. 11[c]: Rendering from the First Position

Element 11[c] recites "simultaneously rendering the two or more of the plurality of digital media streams on the client device starting from the first position by using the descriptor file to synchronize the two or more of the plurality of digital media streams."

DTB discloses simultaneously rendering the two or more of the plurality of digital media streams (e.g., audio and text) on the client device (e.g., player) starting from the first position (e.g., bookmark or lastmark position) by using the descriptor file (e.g., SMIL file) to synchronize the two or more of the plurality of digital media streams (e.g., audio and text). (*Supra* §§VI.A.12 (SMIL file allows synchronized

rendering of audio and text on client device), VI.K.2, VI.C (SMIL file is used to synchronize plurality of digital media streams); EX-1002 ¶¶126-28.)

L. Claim 12

Claim 12 depends from claim 11 and further recites "wherein determining the first position comprises detecting a user interaction with a media rendering utility or a user interface on the client device."

DTB users can "step through and choose from a group of" bookmarks to initiate playback from (and therefore determine) a particular location. (EX-1003, 45-50; EX-1002 ¶130.) Thus, the player's determination of the first position (e.g., bookmark/lastmark position) comprises detecting a user interaction (e.g., choosing bookmark/lastmark) with a media rendering utility or a user interface on the client device (DTB player). (EX-1002 ¶¶129-31.)

M. Claim 13

Claim 13 depends from claim 11 and further recites "wherein determining the first position comprises determining the first position in dependence on the synchronization time offsets in the descriptor file and in dependence on a bookmarked position or a default position."

DTB discloses this limitation. (*Supra* §§VI.D, VI.K; EX-1002 ¶¶132-33.)

N. Claim 14

Claim 14 depends from claim 1 and further recites "wherein the descriptor file contains information that is human readable when rendered."

To the extent claim 14 requires simply a human readable output after the descriptor file has been rendered, DTB's SMIL file contains information (e.g., URIs of text files) that is human readable (text) when rendered. (*Supra* §VI.A.5; EX-1002 ¶135.)

To the extent claim 14 requires the descriptor file to contain human readable text both before and after rendering, this claim is disclosed or rendered obvious by DTB.

As illustrated below, DTB's SMIL file includes human-readable words and numbers:

```
<par id="para1">
   <text region="text" src="rs.xml#para_1" />
   <audio src="rs_fwdx.mp3" clipBegin="00:00:03.51"
   clipEnd="00:01:45.36" />
</par>
```

(EX-1003, 32.) DTB also explains that the SMIL file may contain "one page number as it appears from the print document[.]" (*Id.*, 95-96, 31.) Thus, DTB discloses or renders obvious that the descriptor file (e.g., SMIL file) contains information (e.g., text of SMIL file or page numbers) that is human readable when rendered. (EX-1002 ¶¶134-37.)

O. Claim 17

Claim 17 depends from claim 1 and further recites "wherein the digital media content of each of the plurality of digital media streams includes an audio recording, a video recording, an audio/video recording, a static image, a moving image, an animation, an illustration, text content, footnote content, quotation content, user-entered content, advertising content, or a combination thereof."

DTB discloses this limitation. (*Supra* §VI.A.3; EX-1002 ¶¶138-39.)

P. Claim 18

Claim 18 depends from claim 1.

1. 18[a][i]: Navigating

Element 18[a][i] recites "navigating one or more of the plurality of digital media streams by using the descriptor file on the client device."

DTB discloses "local navigation" that "can be controlled by the ... SMIL file(s)." (EX-1003, 2, 21, 41-50, 7 (DTB players use the synchronization information in the SMIL file to "access points in the audio presentation" and "navigation … is enabled through the textual content file or SMIL file(s)"); *see also supra* §§VI.D, VI.E; EX-1002 ¶141.) Thus, DTB discloses or renders obvious navigating (local navigation or bookmarks/lastmarks) one or more of the plurality of digital media streams (e.g., text or audio streams) by using the descriptor file (e.g., SMIL file). (EX-1002 ¶141.) It also would have been obvious to a POSITA to store

the SMIL file on the client device for "easy access by the player[.]" (*See* EX-1003, 15, 55; EX-1002 ¶¶140-41.)

2. 18[a][ii]: Some Content Not Resident on Client Device

Element 18[a][ii] recites "wherein at least some of the digital media content of the one or more of the plurality of digital media streams is not resident on the client device."

DTB discloses playback of multimedia presentations "over the Internet," which a POSITA would have understood would involve retrieving content that was not originally resident on the client device. (EX-1003, 21; EX-1002 ¶143.)

Even if DTB did not disclose this limitation, it would have been obvious to a POSITA. (EX-1002 ¶144.) A POSITA would have understood content was either stored locally or remotely, both of which would have been obvious. (*Id.*); *KSR*, 550 U.S. at 421. Moreover, a POSITA would have been motivated to enable a client device to download content to enable the user to access a wide range of content without taking up the device's limited storage space. (*Id.*)

Q. Claim 19

Claim 19 depends from claim 1 and further recites "wherein one or more of the plurality of digital media streams are discontinuous media streams that are discontinuous when rendered."

DTB discloses digital talking books with discontinuous media streams (e.g., media streams comprising multiple files). (EX-1002 ¶149.) For example, DTB's text stream can be discontinuous. (EX-1003, 17 ("a DTB can contain multiple textual content files"); EX-1002 ¶149.) Similarly, DTB discloses discontinuous audio streams. (EX-1003, 2, 15 (manifest illustrating multiple audio files); *see also id.*, 61.) Thus, DTB discloses or renders obvious a discontinuous digital media stream (e.g., text and/or audio stream) that is discontinuous when rendered. (EX-1002 ¶¶147-50.)

R. Claim 21

Claim 21 recites "[a] non-transitory computer-readable storage medium having stored thereon instructions which, when executed by a processor of a computing device, cause the computing device to" perform the steps of claim 1.

DTB producers use computers with PC-based browsers and "software" to "generate the SMIL file." (*See* EX-1003, 30-31, 3, 60, 15, 55-56; EX-1002 ¶152.) Thus, DTB discloses or renders obvious claim 21's preamble. (*Id.*)

After the preamble, claim 21 recites the same steps as claim 1. (EX-1002 ¶153.) DTB therefore discloses or renders obvious claim 21. (*Supra* §VI.A; EX-1002 ¶¶151-54.)

VII. GROUND 1B: CLAIMS 2 AND 18-19 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB AND YOSHIMURA.

DTB discloses or renders obvious claims 2, 18, and 19.[2] (*Supra* §VI.) These claims were also obvious in further view of Yoshimura. (EX-1002 ¶¶155-78.)

A. Claim 2

Yoshimura discloses storing SMIL files on a server (e.g., portal server) that is accessible to the client device (e.g., mobile client). (EX-1006, 1785 ("mobile clients simply follow the SMIL file downloaded from the portal server"), 1779 ("Mobile clients … download the modified SMIL files from a portal server"), 1779 (describing benefits of Yoshimura's method), 1781-82; EX-1002 ¶157.)

A POSITA would have been motivated to store DTB's SMIL file on a server, as taught by Yoshimura, for the reasons discussed above. (*Supra* §VI.B; EX-1002 ¶159.)

Moreover, both references contemplate delivery of synchronized audio and text data over a network using SMIL files. (EX-1003, 21; EX-1006, 1779-80; EX-1002 ¶160.)

The combination represents merely the simple addition of one known element (e.g., Yoshimura's servers) with other known elements (e.g., DTB's client device and SMIL file) to obtain predictable results (e.g., a SMIL file stored on a server

[2] Because the claim language is quoted above, it is not repeated here.

accessible to the client device). (EX-1002 ¶161); *KSR*, 550 U.S. at 417.

The combination represents using a known technique (e.g., storing SMIL files on servers) to improve a similar device and method (e.g., DTB's "media units" (hard drives)) in the same way. (*Id.*)

The combination further applies a known technique (e.g., storing SMIL files on servers) to a known device and method (e.g., DTB's SMIL files) that is ready for improvement and yields predictable results (e.g., DTB's SMIL file stored on a server). (*Id.*)

A POSITA would have reasonably expected success for the reasons discussed above, and because both references disclose using the same file type (SMIL file) for the same purpose (synchronization of audio and text data). (*Supra* §VI.B; EX-1003, 21; EX-1006, 1779-80; EX-1002 ¶162.)

Thus, DTB and Yoshimura render claim 2 obvious. (EX-1002 ¶¶156-63.)

B. Claim 18

1. 18[a][i]: Navigating

Yoshimura discloses that the client downloads SMIL files from a server and therefore stores them on the client once downloaded. (EX-1006, 1779, 1782, 1785; EX-1002 ¶164.)

A POSITA would have been motivated to store the descriptor file (e.g., SMIL file) on the device as disclosed in Yoshimura for the reasons discussed above.

(*Supra* §VI.P; EX-1002 ¶166.)

Moreover, the combination represents the simple addition of one known element (e.g., storing a SMIL file locally) to another known element (DTB's player) to obtain a predictable result (a player that stores a SMIL file locally). (EX-1002 ¶167); *KSR*, 550 U.S. at 417.

The combination represents the use of a known technique (e.g., Yoshimura's transfer of the SMIL file to the client) to improve a similar method (DTB's navigation of media streams using a SMIL file) in the same way. (*Id.*)

The combination applies a known technique (e.g., Yoshimura's transfer of the SMIL file to the client) to a known method (e.g., DTB's navigation of media streams using a SMIL file) that is ready for improvement and yields predictable results (e.g., local access to the SMIL file). (*Id.*)

Finally, methods involving transferring descriptor files, including SMIL files, to a client were commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶168 (citing EX-1086, 3:19-31; EX-1029, 3; EX-1106, 888-89).)

2. 18[a][ii]: Some Content Not Resident on Client Device

Yoshimura's clients use information in SMIL files to "receive multimedia content from the best cache servers" (e.g., servers located "close to clients") and to deliver the requested "audio and video payloads to the mobile client." (EX-1006,

1778-79, 1782.) Thus, Yoshimura discloses navigating one or more of the plurality of digital media streams (e.g., audio and video content) by using the descriptor file (e.g., SMIL file) on the client device (e.g., mobile client), wherein at least some of the digital media content of the one or more of the plurality of digital media streams is not resident on the client device (e.g., is stored on cache servers). (EX-1002 ¶170.)

A POSITA would have been motivated to retrieve content from a server for the reasons discussed above. (*Supra* §§VII.A, VII.B.1, VI.P.2; EX-1002 ¶172.)

Moreover, the combination represents the simple addition of one known element (e.g., downloading content from a server as in Yoshimura) to another known element (e.g., DTB's player) to obtain a predictable result (a player that can download content remotely). (EX-1002 ¶173); *KSR*, 550 U.S. at 417.

The combination uses a known technique (e.g., retrieving content from Yoshimura's servers) to improve a similar method (e.g., DTB's playback of a multimedia presentation) in the same way. (*Id.*)

The combination applies a known technique (e.g., Yoshimura's retrieval of content from a remote site) to a known method (e.g., DTB's playback of a multimedia presentation) that is ready for improvement and yields predictable results (e.g., retrieval of content from a remote source). (*Id.*)

The combination also represents one of only two options – either the content is resident on the client device or it is not – both of which would have been obvious

to a POSITA. (EX-1002 ¶174); *KSR*, 550 U.S. at 421.

Finally, a POSITA would have reasonably expected success in this combination because retrieving content from a remote source was commonplace and disclosed by numerous references. (EX-1002 ¶175 (citing EX-1086, 3:19-4:2; EX-1032, 17:4-13; EX-1009; EX-1029, 4).)

Thus, DTB and Yoshimura render claim 18 obvious. (EX-1002 ¶¶164-76.)

C. Claim 19

Yoshimura discloses discontinuous audio and video streams wherein the individual files are fetched (and rendered) separately. (EX-1006, 1780-82; *supra* §§VII.A-VII.B; EX-1002 ¶177.) It would have been obvious to combine DTB and Yoshimura to achieve the claimed results. (*Supra* §§VI.Q, VII.A-VII.B; EX-1002 ¶177.)

Thus, DTB and Yoshimura render claim 19 obvious. (EX-1002 ¶¶177-78.)

VIII. GROUND 1C: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB AND DUNCAN.

DTB discloses or renders obvious claim 7. (*Supra* §VI.) Claim 7 was also obvious in further view of Duncan. (EX-1002 ¶¶179-87.)

DTB discloses or renders obvious element 7[a]. (*Supra* §VI.G.1; EX-1002 ¶179.) Regarding element 7[b], Duncan discloses a digital talking book. (EX-1027, Abstract, ¶[0033]). When Duncan's system "reaches the end of the text of the

current page, it automatically advances the viewer to the next page (loading the next chapter, if necessary) and continues speaking." (*Id.*; EX-1002 ¶181.)

It would have been obvious to a POSITA that DTB's playback systems could be modified to automatically advance the viewer to the next page of text while audio playback occurs, as taught by Duncan (e.g., provide automated page turns as the two or more of the plurality of digital media streams are simultaneously rendered). (EX-1002 ¶182.)

A POSITA would have been motivated to combine Duncan's automatic page turning with DTB's player for several reasons.

First, both references seek to "accommodate sight-impaired readers." (EX-1027 ¶[0004]; EX-1003, vii.) It would have been obvious to a POSITA that Duncan's automatic page turning would have made DTB's system more accessible to print-disabled persons, which was a primary motivation of the DTB standard. (EX-1003, Abstract; EX-1027 ¶¶[0004], [0042]; EX-1002 ¶184.)

Second, the combination represents the simple addition of one known element (e.g., Duncan's automatic page turning) to another known element (e.g., DTB's player) to obtain a predictable result (e.g., a player that automatically turns pages). (EX-1002 ¶185); *KSR*, 550 U.S. at 417.

Third, the combination uses a known technique (e.g., Duncan's automatic page turning) to improve a similar method (e.g., DTB's rendering of a digital talking

book) in the same way. (*Id.*) For example, Duncan's automatic page turn feature is accomplished by processing character offsets in an XML text stream (EX-1027 ¶¶[0041]-[0042]), which DTB also discloses. (EX-1003, 6 (textual content file is an XML file), 45 (disclosing ability to mark a character offset in a textual content file).) The combination would improve the user experience because it would eliminate any need to initiate a page turn. (*Id.*)

Fourth, the combination applies a known technique (e.g., Duncan's automatic page turning) to a known method (e.g., rendering DTB's digital talking book) that is ready for improvement and yields predictable results (e.g., DTB's method utilizing Duncan's automatic page turning feature). (EX-1002 ¶185); *KSR*, 550 U.S. at 417.

Finally, methods to automate page turning were commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶186 (citing EX-1059, 49:6-7; EX-1057, Abstract; EX-1058, 4:32-40).)

Thus, DTB and Duncan render claim 7 obvious. (EX-1002 ¶¶179-87.)

IX. GROUND 1D: CLAIM 8 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB AND HECKERMAN.

DTB discloses or renders obvious claim 8. (*Supra* §VI.) Claim 8 was also obvious in further view of Heckerman. (EX-1002 ¶¶188-96.)

Heckerman discloses a method "for synchronizing audio and text data" representing "the same work" using speech recognition, where silence is modeled as a

word that can be recognized. (EX-1030, Abstract, 4:1-13, Figs. 5-7, 11:49-67.) Heckerman then uses the time stamps associated with the confirmed silences in the audio stream to generate synchronized text/audio files. (*Id.*; EX-1002 ¶190.) Thus, Heckerman discloses selection of synchronization time offsets (e.g., time stamps) in dependence on natural language gaps (e.g., silence), inter-word gaps (e.g., ends of sentences or paragraphs), punctuation marks, or parts of speech. (EX-1002 ¶190.)

A POSITA would have been motivated to use natural language gaps (e.g., silences) as disclosed by Heckerman to determine DTB's synchronization time offsets (e.g., clip begin/end in the parallels) for multiple reasons.

First, a POSITA would have understood that synchronizing at natural language gaps, inter-word gaps, punctuation marks, or parts of speech would lead to a smoother audio rendering. (EX-1002 ¶192.)

Second, a POSITA would have been "motivated to do so in order to exploit acoustic cues to segment recorded speech into semantically meaningful chunks." (EX-1074; EX-1002 ¶193.)

Third, the combination represents the simple addition of one known element (e.g., Heckerman's synchronization offsets) to another known element (e.g., DTB's system) to obtain a predictable result (e.g., synchronizing in dependence on natural language or inter-word gaps). (EX-1002 ¶194); *KSR*, 550 U.S. at 417.

Fourth, the combination uses a known technique (e.g., Heckerman's synchronization offsets) to improve a similar method (e.g., rendering DTB's digital talking book) in the same way. (*Id.*) For example, Heckerman's method generates synchronized audio and text data from the same literary work, which is a stated goal of DTB. (EX-1030, 3:6-10; EX-1003, 21 (describing use of SMIL files to synchronize audio and text to create a digital talking book); EX-1002 ¶194.)

Fifth, the combination applies a known technique (e.g., Heckerman's synchronization offsets) to a known method (e.g., DTB's rendering of digital talking books) that is ready for improvement and yields predictable results (e.g., smoother audio rendering). (EX-1002 ¶194); *KSR*, 550 U.S. at 417.

Finally, according to a European Examiner evaluating a similar claim in a foreign equivalent of the '907 patent, segmenting audio based on natural language gaps was "merely one of several alternative straightforward possibilities which the skilled person would select[,]" "notoriously well-known" (EX-1072, 5), and disclosed by numerous references. A POSITA would have had a reasonable expectation of success in the modification. (EX-1002 ¶195 (citing EX-1069, 2:4-6; EX-1063, 113-15; EX-1075, 9-10).)

Thus, DTB and Heckerman render claim 8 obvious. (EX-1002 ¶¶188-96.)

X. GROUND 1E: CLAIMS 14 AND 20 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB AND BULTERMAN.

DTB discloses or renders obvious claim 14. (*Supra* §VI.) Claims 14 and 20 were also obvious in further view of Bulterman. (EX-1002 ¶¶197-209.)

A. Claim 14

Bulterman teaches embedding text directly in a SMIL file. (EX-1007, 86-87). Specifically, Example 5-1(c) explains that instead of referencing an external file (e.g., "externalfile.txt") for a text source, one can embed text directly into a SMIL file by using the "src" attribute "data:". (*Id.*, 86.)

```
1 <smil>
2  <head>
3   ...
4  </head>
5  <body>
6   ...
7   <text src="data:,Greetings!" type="text/plain" ... />
8   ...
9  </body>
10 </smil>
```

(c) using <text> and an embedded media object

Example 5-1. Partitioning control and content in the SMIL Language profile.

(*Id.*) Bulterman discloses that "[f]or short text strings, this is a convenient mechanism[.]" (*Id.*, 87.) Thus, Bulterman discloses this limitation. (EX-1002 ¶199.) This was also disclosed in many other references. (*See, e.g.*, EX-1101, 2 ("literal text" can be included in the SMIL file using a special form of the "src" attribute); EX-1103, 6-7 (SMIL uses "in-line text instead of referring to seperate [sic] plain-text files as the text source."); EX-1002 ¶200.)

A POSITA would have been motivated to directly embed text in DTB's SMIL file for several reasons.

First, a POSITA would have understood that eliminating the need to reference or point to a text file (as was done in DTB) and instead include actual text (especially for short text passages) would lead to faster processing times and would lower or eliminate the likelihood of service failure due to slow or failed network connections in the case where text files were saved on and retrieved from servers. (EX-1002 ¶202.) Indeed, Bulterman advises that "[f]or short text strings, this is a convenient mechanism[.]" (EX-1007, 87.)

Second, a POSITA would have understood that a limited number of options existed for how a SMIL file identifies text, i.e., using a text file located outside of the SMIL file or including the text within the SMIL file. (EX-1002 ¶203.) Both would have been obvious to a POSITA and a POSITA would immediately have understood Bulterman's inclusion of text in the SMIL file to be a viable method. (*Id*.)

Third, the combination represents merely the simple addition of one known element (e.g., Bulterman's text included directly in the SMIL file) to another known element (e.g., DTB's SMIL file) to obtain predictable results (e.g., a more efficient SMIL file). (EX-1002 ¶204); *KSR*, 550 U.S. at 417.

Fourth, the combination uses a known technique (e.g., Bulterman's inclusion of text to be rendered in the SMIL file) to improve a similar method (e.g., DTB's

SMIL file) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., inclusion of the text to be rendered within the SMIL file) to a known method (e.g., DTB's method SMIL file) that is ready for improvement and yields predictable results (e.g., a SMIL file with less likelihood for delay or broken links). (*Id.*)

Finally, including text to be rendered within the SMIL file was commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶205 (citing EX-1101, 2 ("literal text" can be "included in the SMIL" file using a special form of the "src" attribute); EX-1103, 6-7; EX-1051, 114; EX-1104, 83.)

Thus, DTB and Bulterman render claim 14 obvious. (EX-1002 ¶¶198-206.)

B. Claim 20

Claim 20 depends from claim 1 and further recites "wherein the location information for one or more of the plurality of digital media streams indicates that the one or more of the plurality of digital media streams are located within the descriptor file."

The "src" attribute specifies the storage location of digital media streams. (*Supra* §VI.A.5; EX-1003, 98 (the "src" attribute "specifies by URI the location of the image file"); EX-1002 ¶208.) Bulterman's "src" attribute indicates the actual text is included in the SMIL file by including the characters "data:" before the actual

content. (EX-1007, 86-87.) Thus, DTB and Bulterman render claim 20 obvious. (*Supra* §X.A; EX-1002 ¶¶207-09.)

XI. GROUND 1F: CLAIM 15 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB AND YANG.

Claim 15 depends from claim 1 and further recites "coordinating delivery of the digital media content of the plurality of digital media streams from one or more servers to the client device in dependence on the synchronization time offsets in the descriptor file."

Yang discloses this limitation. (EX-1002 ¶¶210-11.) For example, Yang discloses a "just-in-time" retrieval system for synchronized multimedia presentations using SMIL files. (EX-1045, 49.) Yang's "object-retrieving engine first parses the input SMIL document to extract and to represent the synchronization relationship of the objects in the presentation[.]" (*Id.*, 51.) "Next, the object request time for each object is determined by considering the user interaction [and] the playback time of the object[.]" (*Id.*) "The request for retrieving the object is then issued by the object-retrieving engine at the object request time." (*Id.*)

A POSITA would have been motivated to deliver DTB's content with Yang's just-in-time method for several reasons.

First, Yang explains that its "just-in-time retrieving policy" is feasible and has better performance as proven "by performance measurements of system implementation." (EX-1045, 49; EX-1002 ¶213.)

Second, the combination represents merely the simple addition of one known element (e.g., Yang's just-in-time retrieval based on SMIL timing information) to another known element (e.g., DTB's player) to obtain predictable results (e.g., prefetching content). (EX-1002 ¶214); *KSR*, 550 U.S. at 417.

Third, the combination uses a known technique (e.g., Yang's just-in-time retrieval) to improve a similar method (e.g., DTB's retrieval of information identified in a SMIL file) in the same way. (*Id.*) The combination would improve the user experience because it would result in a smoother streaming session. (EX-1006, 1782; EX-1002 ¶214.)

Fourth, the combination applies a known technique (e.g., Yang's object retrieval) to a known method (DTB's delivery of content) that is ready for improvement and yields predictable results (e.g., prefetched objects). (EX-1002 ¶214); *KSR*, 550 U.S. at 417.

Finally, methods involving prefetching segments were commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶215 (citing EX-1006, 1782-83; EX-1086, 4:8-25; EX-1105, 577).)

Thus, DTB and Yang render claim 15 obvious. (*Supra* §V1.A; EX-1002 ¶¶210-16.)

XII. GROUND 1G: CLAIM 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, YANG, AND COPLEY.

Claim 16 depends from claim 15 and further recites "managing network connections between the client device and the one or more servers in dependence on service failure, service degradation, digital data throughput rate, or a combination thereof."

Copley discloses a streaming media system that uses "client-side performance monitoring to automatically detect and correct connection failures or performance problems." (EX-1009 ¶[0016].) "When such difficulties are encountered, [the system] automatically switches the streaming activity to the next optimum provider to minimize disruption to the media delivery." (*Id.*) The monitored performance problems are analyzed based on performance data including failure information, latency information, and transport activity. (*Id.*, claim 18, ¶¶[0028], [0032], [0047]-[0048], [0065], [0074]-[0075].) Copley also discloses managing network connections in dependence on service degradation (e.g., buffer stall, playback stall, unacceptable packet loss, etc.). (*Id.* ¶¶[0086]-[0089]; EX-1002 ¶218.)

A POSITA would have been motivated to combine the teachings of DTB, Yang, and Copley, and to manage network connections based on service failure, service degradation, and/or digital data throughput rate as described in Copley for several reasons.

First, Copley explains that doing so provides "a high quality of service," "low development costs, and rapid time to market for both turnkey and customized media delivery solutions." (EX-1009 ¶¶[0013]-[0015]; EX-1002 ¶220.)

Second, the combination represents merely the simple addition of one known element (e.g., Copley's server optimization) to another known element (e.g., DTB's player as modified by Yang) to obtain predictable results (e.g., presentation of multimedia via multiple servers). (EX-1002 ¶221); *KSR*, 550 U.S. at 417.

Third, the combination uses a known technique (e.g., Copley's server optimization) to improve a similar method (e.g., DTB's presentation of multimedia as modified by Yang) in the same way. (*Id.*) The combination would improve the user experience because the streaming content would suffer fewer interruptions. (EX-1002 ¶221.)

Fourth, the combination applies a known technique (e.g., Copley's server optimization) to a known method (e.g., DTB's presentation of multimedia) that is ready for improvement and yields predictable results (e.g., presentation of multimedia via multiple servers). (EX-1002 ¶221); *KSR*, 550 U.S. at 417.

Finally, server optimization based on failure conditions, network degradation, and throughput was commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶222 (citing EX-1032, 12:53-67; EX-1029, 4-5; EX-1006, 1781-82).)

Thus, DTB, Yang, and Copley render claim 16 obvious. (*Supra* §§VI.A, XI; EX-1002 ¶¶217-23.)

XIII. GROUND 2A: CLAIMS 1-3, 8-12, 14, 17-19, AND 21 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY.

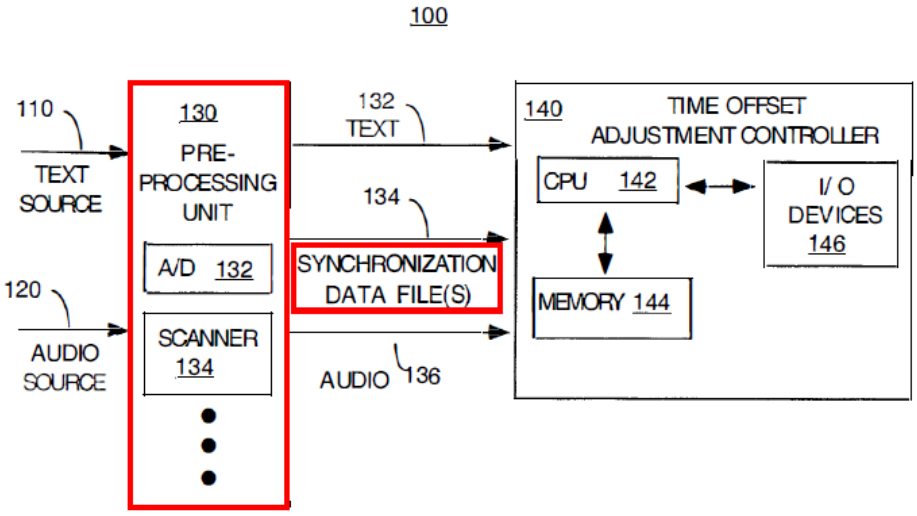
A. Claim 1

1. Preamble

McCartney discloses or at least renders the preamble obvious because it discloses a "[m]ethod for constructing a digital talking book from text data and audio data." (EX-1013, claim 1; EX-1002 ¶225.)

2. 1[a][i]: Descriptor File

McCartney discloses the production and rendering of digital talking books using "synchronization files, e.g., a book project management (BPM) file and a Time Stamp Data (TSD) file" to coordinate text and audio data. (EX-1013, Abstract, Fig. 1, ¶¶[0026], [0031], [0019], [0021].) The BPM file contains "information on which text elements [] are synchronizable and/or navigable." (*Id.* ¶[0026].) The TSD file contains information about time points within each audio recording that "are to be synchronized with specific elements in the marked-up text." (*Id.* ¶[0031].) As illustrated below, these synchronization files are generated by a pre-processing unit. (*Id.* ¶¶[0022], [0038].)



(*Id.*, Fig. 1; EX-1002 ¶227.)

Accordingly, McCartney discloses or renders obvious creating descriptor files (e.g., TSD file and BPM file) for synchronizing a plurality of digital media streams (e.g., text and audio streams). (EX-1002 ¶¶226-28.)

Even if this limitation required a single descriptor file, it would have been obvious to a POSITA in view of McCartney. (EX-1002 ¶229.) It would have been obvious to a POSITA to combine McCartney's BPM and TSD files' data into a single file having characteristics of both the TSD and BPM files. (*Id.*) A POSITA would have been motivated to do so at least because it would make it easier to manage, review, share, access, and revise the information contained in the files. (*Id.*) Moreover, a POSITA would have had a reasonable expectation of success in combining the TSD and BPM files because McCartney explains that both TSD and BPM files can processed by "an XML application," meaning they share a compatible format. (EX-1013 ¶¶[0026], [0031]; EX-1002 ¶230.)

3. 1[a][ii]: Same Originating Work

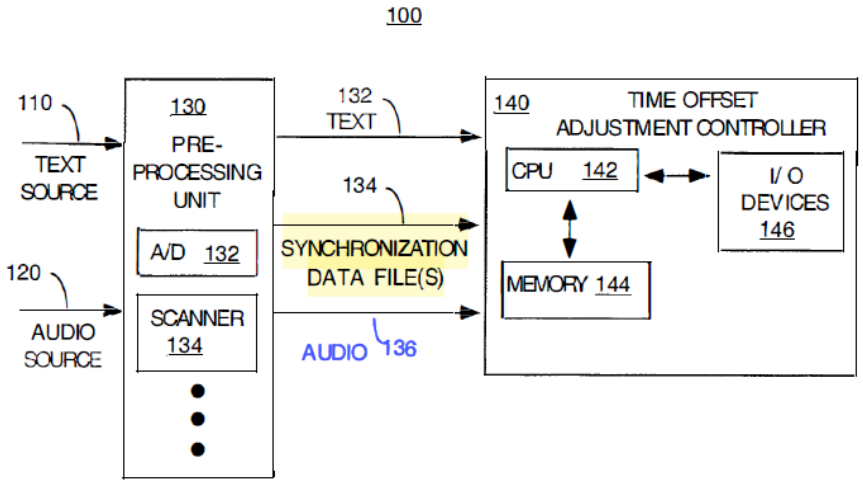
McCartney's text files "contain the full text of the original printed book." (EX-1013 ¶[0020].) McCartney's audio files "contain the full-recorded text of the printed book." (*Id.* ¶[0021].) Thus, McCartney discloses that the plurality of digital media streams (e.g., audio and text streams) each contain digital media content corresponding to a same originating work (e.g., printed book). (EX-1002 ¶¶232-33.)

4. 1[a][iii]: Digital Media Streams

McCartney discloses that the plurality of digital media streams (e.g., audio and text streams) includes a first digital media stream containing a digital audio narration (e.g., audio stream) of the originating work (e.g., printed book) and one or more other digital media streams (e.g., text stream). (EX-1013 ¶¶[0020]-[0021]; *supra* §§XIII.A.2-XIII.A.3; EX-1002 ¶¶234-35.)

5. 1[a][iv]: External Descriptor File

Figure 1 of McCartney shows that the synchronization files (e.g., TSD and BPM files) (highlighted) are external to the first digital media stream (audio stream) (blue):



(EX-1013, Fig. 1, ¶[0031]; *id.* ¶¶[0038]-[0039]; EX-1002 ¶¶236-37.)

6. 1[b]: Location Information

McCartney discloses or renders obvious storing location information in its TSD and BPM files. (EX-1002 ¶¶238-43.)

a. TSD File

The TSD file identifies synchronized audio files by name (e.g., "track6.wav"). The filename is location information because it can be used in locating the file. (EX-1002 ¶239.) For example, a POSITA would have understood that files were located using a file path, part of which is the file name. (*Id.*) Alternatively, it would have been obvious to a POSITA based on McCartney's disclosure that "[v]arious attributes can be employed … for defining the name of the audio file," to include a URL, complete file path, pointer, or some other location information in the <data> element for locating the audio stream. (EX-1013 ¶[0032]; EX-1002 ¶239.) A

POSITA would have been motivated to include location information to assist in locating the referenced audio file. (*Id.*)

Alternatively, to the extent "location information" includes a location within the file (e.g., play location), the TSD file also includes this type of information. For example, the <record> elements reflect play times within the audio file and navigation points to the corresponding text file. (EX-1013 ¶¶[0032]-[0033], [0042]-[0043], Fig. 5; EX-1002 ¶240.)

b. BPM File

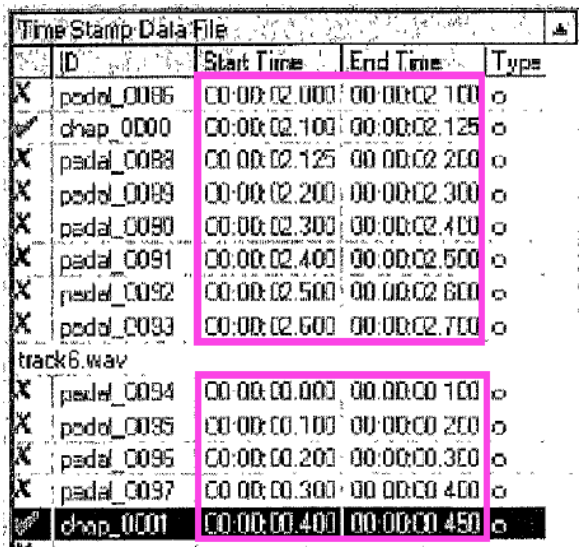
McCartney's BPM file includes "a list of the marked-up text files that make up the book" and a "path" attribute that holds "either a relative or absolute path to the location of the [text] file on the storage media." (*Id.* ¶[0029], 7; EX-1002 ¶241.) Thus, McCartney's BPM file includes file location information (e.g., file names and/or paths to the location of the file) for the text stream. (EX-1002 ¶241.)

Alternatively, to the extent "location information" includes a location within the file, the BPM file also includes this type of information. For example, the BPM file "contains information that identifies synchronizable or navigation elements" (e.g., "headings, pages, etc.") in the text files. (EX-1013 ¶[0030]; EX-1002 ¶242.)

7. 1[c][i]: Time Offsets

Each audio clip in McCartney's audio stream is expressed in the TSD file as a <record> element, containing a unique ID, the clip starting time, the clip ending

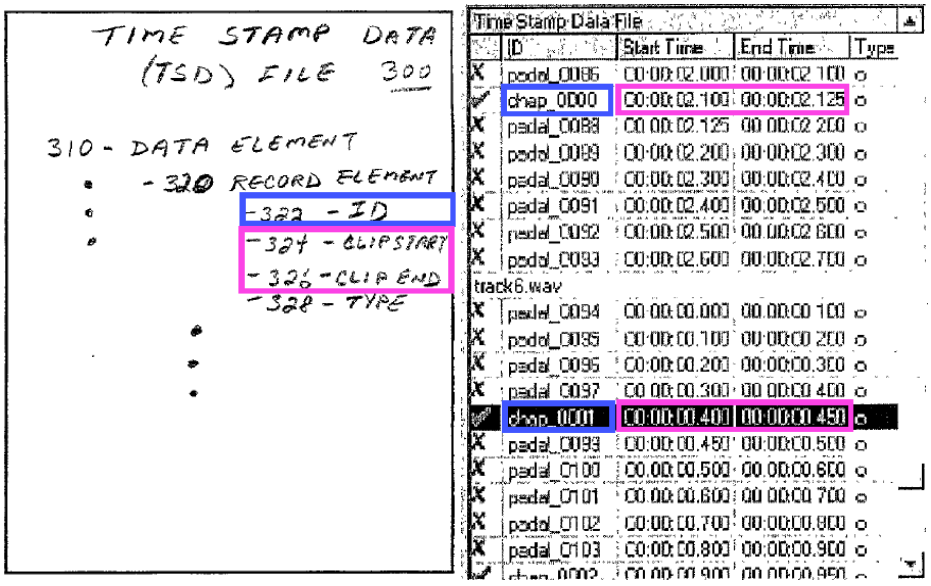
time, and type. (EX-1013 ¶[0033].) This can be seen in Figure 5 below, where the **clip starting time** and **clip ending time** are highlighted in **pink**.



(*Id.*, Fig. 5, ¶[0033]; EX-1002 ¶245.) McCartney's TSD files, including the time offsets, are generated at the pre-processing unit. (EX-1013 ¶[0023].) Thus, McCartney discloses identifying a plurality of time offsets (e.g., start time and/or end times) in a timeline of the digital audio narration (audio file) of the first digital media stream (audio stream). (EX-1002 ¶¶244-46.)

8. 1[c][ii] and 1[d]: Content Points Correlated to Time Offsets

Each of McCartney's <record> elements contains an ID attribute, which identifies the audio clips by their associated content (e.g., by heading, chapter (as illustrated below), or page). (EX-1013 ¶¶[0031]-[0033], [0040].) The correlation between the content point IDs (blue) and start/end times (pink) is stored in the TSD file as shown below:



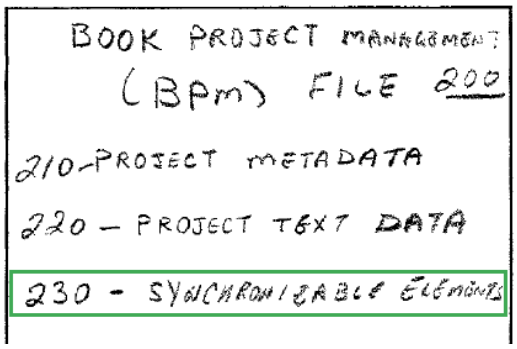
(*Id.*, Figs. 5, 3; EX-1002 ¶248.) Thus, McCartney discloses or renders obvious storing the plurality of time offsets (e.g., clip start and/or end times) and the plurality of content points (e.g., content identified in ID attributes) in the descriptor file (e.g., TSD file) in a manner indicating a correlation (e.g., in the same line) between the plurality of time offsets and the plurality of content points. (EX-1013 ¶[0033], Figs. 3, 5; EX-1002 ¶¶247-49.)

9. 1[e]: Synchronization Points

In McCartney, the "other digital media stream" is text presented, for example, in an HTML file. (EX-1013 ¶¶[0019]-[0020]; *supra* §XIII.A.4.) McCartney links the text and audio files by chapters (sometimes referred to as "headings") and/or pages. (EX-1013 ¶¶[0041]-[0043], [0046].) Each chapter/page is assigned an ID attribute and the text and audio streams are linked by assigning identical ID attributes to corresponding portions of the text and audio stream. (*Id.*) Linked HTML (text)

elements and audio elements (e.g., portions of the audio stream that share the same ID) are identified by a green checkmark in the TSD file. (*Id.*; *supra* §XIII.A.8 (ID attributes correspond to positions in the audio); EX-1002 ¶251.)

Alternatively, McCartney's BPM file lists "synchronizable elements within the source text files that are used to indicate points of navigation." (EX-1013, 7; *see also id.* ¶[0030] (BPM file contains "identifications of which classes of elements [(headings, pages, etc.)] are to be considered points of synchronization").) This is illustrated in Figure 2 of McCartney:



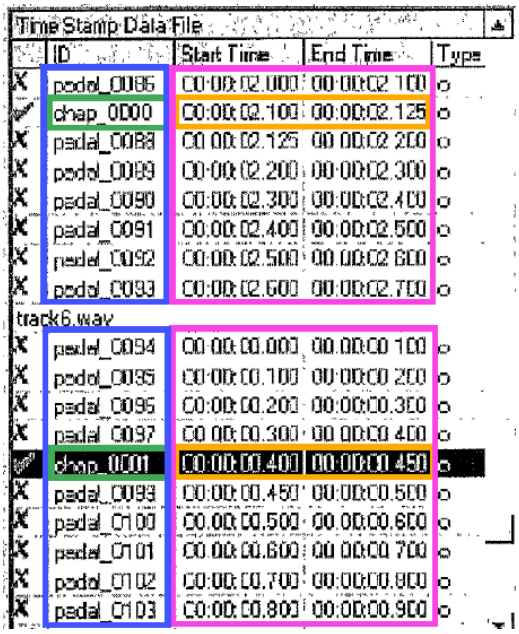
(*Id.*, Fig. 2; EX-1002 ¶252.)

Thus, McCartney discloses identifying synchronization points (e.g., IDs with a green check in the TSD file and/or synchronizable elements in the BPM file) in the digital media content of the one or more other digital media streams (e.g., text files). (EX-1013 ¶¶[0038], [0023]; EX-1002 ¶¶250-53.)

10. 1[f]: Synchronization Time Offsets

When a "point in the source text file" is assigned an ID attribute matching the corresponding content point in the audio stream, that point in the text stream is

"linked to" a time offset. (EX-1013 ¶¶[0033], [0043], [0052].) Linked audio and text events are indicated in the TSD file by a "checkmark," whereas unlinked events are identified by a "X." (*Id.* ¶[0043].) Figure 5 shows a plurality of content points (blue), a plurality of time offsets (pink), linked synchronization points (green), and their corresponding synchronization time offsets (yellow):



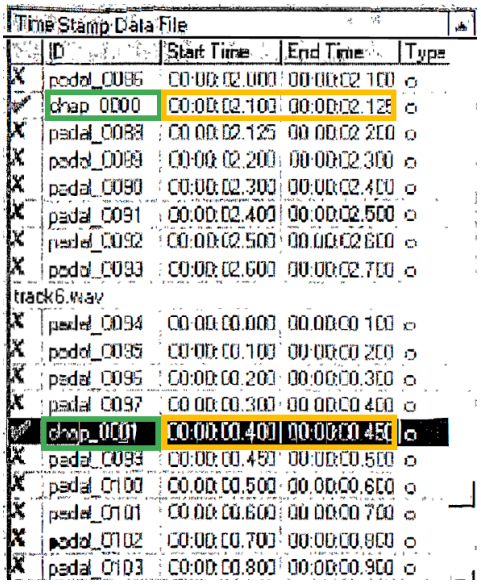
(*Id.*, Fig. 5; EX-1002 ¶255.)

McCartney's synchronization time offsets are selected when text elements are selected to be linked to TSD events (e.g., audio clips) and selected again if the timing of those offsets are adjusted. (EX-1013 ¶¶[0050], [0052].) McCartney's synchronization time offsets are also a subset of the plurality of time offsets and are therefore selected from the plurality of time offsets. (*Id.* ¶¶[0042], [0033]; EX-1002 ¶256.) Thus, McCartney discloses or renders obvious selecting synchronization time offsets

(e.g., clip start and end attributes for linked text and audio events) that correspond to the synchronization points (e.g., points in text stream identified by an ID attribute) from the plurality of time offsets (e.g., clip start and end attributes for the audio stream). (EX-1002 ¶¶254-57.)

11. 1[g][i]: Correlating Synchronization Time Offsets and Synchronization Points

Figure 5 of McCartney illustrates that synchronization time offsets (yellow) and synchronization points (green) are stored in the descriptor file (e.g., TSD file) in a manner indicating a correlation between them (e.g., as part of the same, inline data element):



(EX-1013, Fig. 5; *id.* ¶[0033] (ID attributes correspond to clip start/end times); EX-1002 ¶¶258-60.) McCartney therefore discloses or renders obvious this limitation. (*Id.*)

12. 1[g][ii]: Synchronized Rendering

McCartney discloses a "time offset adjustment controller" (TOAC) that uses the TSD and BPM files to synchronize the audio and text streams to produce a "properly coordinated and constructed digital talking book." (EX-1013 ¶¶[0022], [0046], [0052], [0009].) McCartney's TOAC operator can "click on an event in the TSD list to hear the audio clip represented by [an audio] event while the associated [text] element [] is highlighted." (*Id.* ¶¶[0009], [0044]-[0046], [0024], Figs. 1, 5-6.)

Alternatively, McCartney discloses that the TOAC can "generate SMIL files from the TSD files" and an index (referred to as a Navigation Control Center or NCC) based on the BMP file so that a digital talking book could be rendered on a user's device. (EX-1013 ¶¶[0057]-[0058]; EX-1002 ¶263.) The TSD and BMP files allow synchronized rendering of the text and audio stream because they are used in creating the digital talking book for synchronized rendering on a client device. (*Id.*)

Accordingly, McCartney discloses or renders obvious a descriptor file (e.g., TSD and BPM files) that allows a synchronized rendering of the plurality of digital media streams (e.g., text and audio) on a client device (e.g., TOAC or DTB player). (EX-1002 ¶¶261-64.)

B. Claim 2

McCartney's system comprises: (1) a preprocessing unit for generating the TSD and BPM files; and (2) a TOAC for rendering content using the TSD and BPM

files. (EX-1013, Fig. 1, ¶¶[0022]-[0025]; EX-1002 ¶267.) The preprocessing unit and the TOAC are each implemented on a general purpose computer, which may "include one or more personal computers, servers, main frames and the like." (EX-1013 ¶¶[0025] (TOAC general purpose computer may include a server), [0023] (preprocessing unit's general purpose computer may be "similar to that of the TOAC").)

McCartney's TSD and BPM files are stored, at least temporarily, on the pre-processing unit because the pre-processing unit creates those files. (*Id.*) The files are also stored on, and used by, the TOAC. (*Id.*) It would have been obvious to a POSITA to store the descriptor file on a server (e.g., the pre-processing unit server or the TOAC server) that is accessible to the client device (TOAC). (EX-1002 ¶268.)

A POSITA would have been motivated to store McCartney's TSD and BPM files on a server accessible to TOAC's input/output computer. (*Id.* ¶269.) A POSITA would have understood that storing the TSD and BPM files on a server allows increased accessibility. (*Id.*) For example, multiple TOAC users could access and edit any given TSD/BPM file. (*Id.*) Moreover, any given TOAC user could access the TSD/BPM files from any TOAC system. (*Id.*) A POSITA also would have understood that storing the TSD and BPM files on a server allows for more scalability (e.g., the ability to create and edit more TSD and BPM files for more talking books at any given time). (*Id.*)

Thus, McCartney discloses or renders obvious claim 2. (EX-1002 ¶¶266-71.)

C. Claim 3

McCartney discloses or renders obvious this limitation. (EX-1002 ¶¶272-73; *supra* §XIII.A.12.)

D. Claim 8

McCartney's BPM file identifies which classes of elements (e.g., "headings, pages, etc.") are to be considered "points of synchronization." (EX-1013 ¶¶[0030], [0038], [0052]; EX-1002 ¶275.)

McCartney also incorporates Woodill by reference. (EX-1013 ¶[0022].) Woodill discloses identifying time offsets in audio by detecting periods of silence within the audio file to isolate events in the audio file. (EX-1014 ¶¶[0001], [0003], [0027]-[0029], [0036]-[0038], claim 1; EX-1002 ¶276.) Time offsets are then extracted and identified as corresponding to pages or chapters. (*Id*.) Thus, McCartney discloses (via Woodill), or renders obvious, identifying a plurality of time offsets in dependence on natural language gaps (e.g., silence). (*Id*.; EX-1002 ¶¶274-78.)

E. Claims 9-10

McCartney's audio data "is typically independently-produced audio clips" or files. (EX-1013 ¶¶[0021], [0039].) The time data for each audio file is contained in the TSD file's <data> element, which contains the "clip starting time" and "clip ending time" for each audio file. (*Id.* ¶¶[0032]-[0033], Figs. 3, 5.)

Thus, McCartney discloses or renders obvious a first digital media stream (e.g., audio stream) including a plurality of content segments (e.g., audio clips and/or audio files) defined by the plurality of time offsets (e.g., clip starting and ending times) in the descriptor file (e.g., TSD file). (EX-1002 ¶¶279-82.)

F. Claims 11-12

1. 11[a]: Selecting Streams

McCartney discloses "an operator may select a book section to work with and open it." (EX-1013 ¶¶[0038]-[0039].) Alternatively, a POSITA would have understood a user could select a book using a media player. (EX-1002 ¶284.) In each case, by selecting a book, the operator or user is selecting two or more of the plurality of streams. (*Id.*) Accordingly, McCartney discloses or renders obvious this limitation. (*Id.* ¶¶283-84.)

2. 11[b]: Determining a First Position and 12: User Interaction

McCartney's TOAC operator can select "events" (e.g., text with a synchronized audio clip) by user interaction with the TOAC (e.g., by clicking on an event). (EX-1013 ¶¶[0009], [0046], [0042]-[0043], [0024]; EX-1002 ¶286.) The text and audio clips associated with each event are synchronized by the time offsets of the audio clip (synchronization offsets). (*Id.*) When an event is selected, the associated audio clip is therefore determined by referencing the audio clip offsets. (*Id.*) Similarly, the text segment is determined based on which text passage matches the audio

clip time offsets. (*Id.*) Because the text and audio files are linked by the start and end time of the audio clip associated with that portion (e.g., chapter), the TOAC determines a first position in the plurality of digital media streams in dependence on the synchronization offsets in the TSD file. (*See* EX-1013 ¶¶[0009] (clicking on linked events "caus[es] the TOAC to begin playing the associated audio clip"), [0042].) Accordingly, these limitations were disclosed or rendered obvious by McCartney. (EX-1002 ¶286.)

3. 11[c]: Rendering from the First Position

Clicking on a linked event in the TSD list initiates synchronized playback of the audio and text streams from the point represented by that TSD event. (EX-1013 ¶[0046]; EX-1002 ¶288; *supra* §XIII.A.12.)

McCartney therefore discloses or renders obvious claims 11 and 12. (EX-1002 ¶¶283-89.)

G. Claim 14

To the extent claim 14 requires simply a human readable output after the descriptor file has been rendered, McCartney's TSD and BPM files contain information (e.g., identification of text files) that is human readable (text) when rendered. (*Supra* §XIII.A.2; EX-1002 ¶291.)

To the extent claim 14 requires the descriptor file to contain human readable text both before and after rendering, this is disclosed or rendered obvious by

McCartney. McCartney's TSD file includes human readable text that is not removed when the TSD file is rendered including, for example, comments relating to the audio segment, ID, start time, end time, etc. (EX-1013, Fig. 5, 11-13; EX-1002 ¶292.)

It would have been obvious to include the actual text in the ID attribute of the TSD file and/or the "path" section of the BPM file. (*Supra* §§VI.N, X.A; EX-1013, 7, 12; EX-1002 ¶293.)

Thus, McCartney renders claim 14 obvious. (EX-1002 ¶¶290-95.)

H. Claim 17

McCartney discloses the limitations of claim 17. (*Supra* §XIII.A; EX-1002 ¶¶296-97.)

I. Claim 18

McCartney's user navigates (e.g., by clicking or selecting the book section and/or editing book section links) one or more of the plurality of digital media streams (e.g., audio and/or text) by using the descriptor file (e.g., TSD) on the client device (e.g., TOAC). (*Supra* §§XIII.A.12, XIII.F; EX-1002 ¶299.)

Storing media content such that it was not resident on the TOAC would have been obvious in view of McCartney. (*Supra* §§VI.P, XIII.B; EX-1002 ¶300.)

Thus, McCartney renders claim 18 obvious. (EX-1002 ¶¶298-303.)

J. Claim 19

McCartney discloses rendering discontinuous audio and text streams. (EX-1013 ¶¶[0032], [0023], 7-9 (indicating multiple source text files for a given book (e.g., book project)), Fig. 3; EX-1002 ¶305.) McCartney never discloses reassembling the separate/discontinuous files into a single file prior to rendering. (EX-1002 ¶305.) McCartney therefore discloses or renders obvious claim 19. (EX-1002 ¶¶304-06.)

K. Claim 21

McCartney discloses or renders obvious the limitations of claim 21. (EX-1002 ¶¶307-10; *supra* §§VI.R, XIII.A; EX-1013 ¶¶[0022]-[0025].)

XIV. GROUND 2B: CLAIMS 2 AND 18 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND KAUFFMAN.

McCartney discloses or renders obvious claims 2 and 18. (*Supra* §XIII.) They were also obvious in further view of Kauffman. (EX-1002 ¶¶311-21.)

A. Claim 2

Kauffman discloses an object server storing a media object divided into a plurality of pieces and a "piece map." (EX-1015, 6:40-57.) Kauffman's "piece map contains an ordered sequence of identifiers and lengths of the actual content objects or pieces." (*Id*., 7:61-63.) Kauffman further discloses that the piece map is accessible to the client. (*Id*., 11:3-4.) Thus, Kauffman discloses storing a descriptor file

(e.g., piece map) on a server (e.g., object server) that is accessible to the client device (e.g., library client). (EX-1002 ¶313.)

A POSITA would have been motivated to store McCartney's TSD and BPM files on a server accessible to TOAC's input/output computer for the reasons discussed above. (*Supra* §XIII.B; EX-1002 ¶315.)

Moreover, the combination represents merely the simple addition of one known element (e.g., server storage) to another known element (e.g., McCartney's system) to obtain predictable results (e.g., descriptor files stored on a server). (EX-1002 ¶316); *KSR*, 550 U.S. at 417.

The combination uses a known technique (e.g., server storage) to improve a similar method (e.g., McCartney's computer storage) in the same way. (*Id.*)

The combination applies a known technique (e.g., server storage) to a known method (e.g., McCartney's computer storage) that is ready for improvement and yields predictable results (e.g., descriptor files saved on a server). (*Id.*)

Finally, storing descriptor files on servers was commonplace and disclosed in other references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶317 (citing EX-1106, 889-91; EX-1006).)

Thus, McCartney and Kauffman render claim 2 obvious. (EX-1002 ¶¶312-18.)

B. Claim 18

A POSITA would have been motivated to store content remotely for the reasons discussed above (*supra* §§ XIII.I, XIV.A) and because it represents nothing more than the simple use of a known element (remote storage of media) within a known system (McCartney's) to obtain predictable results (content storage and retrieval). (EX-1002 ¶320;) *KSR*, 550 U.S. at 417. A POSITA would have reasonably expected success in making the modification because storing and transferring media streams from remote servers was well known and conventional. (EX-1002 ¶320 (citing EX-1029, 2-3, Fig. 2; EX-1106, 888-89).)

Thus, McCartney and Kauffman render claim 18 obvious. (EX-1002 ¶¶319-21.)

XV. GROUND 2C: CLAIMS 4-6 AND 13 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND DTB.

A. Claim 4

1. 4[a]: Tracking Position

McCartney's TOAC tracks the position of the audio and text streams as the user checks the accuracy of the links between the two streams. (EX-1013 ¶¶[0037]-[0056]; EX-1002 ¶324.) McCartney's system pre-populates links between sections (e.g., chapters) of the audio and text streams. (EX-1013 ¶¶[0041]-[0044]; EX-1002 ¶324.) McCartney then asks a user whether they want to check the accuracy of each link. (EX-1013 ¶[0045]; EX-1002 ¶324.) After the accuracy of the link is checked

by rendering both the audio and text for that section, McCartney advances to the next section and asks the user whether they want to check the accuracy of the link for that section. (EX-1013 ¶¶[0044]-[0046], [0056]-[0058], Fig. 6; EX-1002 ¶324.) McCartney therefore discloses or renders obvious this limitation. (EX-1002 ¶¶323-24.)

Even if McCartney did not disclose this limitation, it would have been obvious in view of DTB. (*Supra* §VI.D; EX-1002 ¶325.)

A POSITA would have been motivated to track a current position in McCartney's media streams for multiple reasons.

First, it would assist the TOAC operator in verifying that each section of text and audio elements have been correctly linked. (EX-1013 ¶¶[0046], [0052] (TOAC operator verifies text/audio links using time information); EX-1002 ¶326.)

Second, the combination represents merely the simple addition of one known element (e.g., DTB's progress tracking/markings) to another known element (e.g., McCartney's system) to obtain predictable results (e.g., tracking progress while rendering content on TOAC). (EX-1002 ¶327); *KSR*, 550 U.S. at 417.

Third, the combination uses a known technique (e.g., DTB's progress tracking/marking) to improve a similar method (e.g., McCartney's TOAC rendering) in the same way. (*Id.*)

Fourth, the combination applies a known technique (e.g., DTB's

tracking/marking position) to a known method (e.g., McCartney's content rendering) that is ready for improvement and yields predictable results (e.g., tracking progress when rendering content on TOAC). (*Id.*)

A POSITA would have had a reasonable expectation of success in implementing the tracking because McCartney and DTB are related to similar systems using similar formats (e.g., XML/SMIL) and because tracking a position in a media stream was well known in the art. (EX-1033, 5:4-6 ("the system monitors the time ... that the user has progressed into the audio"); EX-1002 ¶328.)

2. 4[b]: Bookmark

DTB discloses this limitation. (*Supra* §VI.D; EX-1002 ¶329.)

A POSITA would have been motivated to modify McCartney so that the TOAC operator can create a bookmark that includes the bookmarked position (e.g., ID or time) and identifies the descriptor file because doing so would enable the TOAC operator to stop and then resume editing at the identified position of that file. (EX-1002 ¶330.) For example, McCartney's TOAC "can operate on one 'book section' at a time." (EX-1013 ¶[0039].) If a user wishes to move to another section of the book, or a different book altogether, a POSITA would have understood the benefit of marking where the user left off in one book section (e.g., by saving the identity of the TSD file and position within the TSD file). (EX-1002 ¶330.)

A POSITA would have been motivated to create a bookmark and would have

had a reasonable expectation of success in doing so in McCartney's system. (EX-1002 ¶331; *see also supra* §XV.A.1.)

Thus, McCartney and DTB render claim 4 obvious. (*Supra* §XIII; EX-1002 ¶¶323-32.)

B. Claim 5

DTB discloses claim 5. (*Supra* §VI.E; EX-1002 ¶334.)

It would have been obvious to a POSITA to use the bookmark to render the content on the same or a different TOAC device. (*See supra* §XV.A; EX-1002 ¶335.)

Thus, McCartney and DTB render claim 5 obvious. (*Supra* §XIII; EX-1002 ¶¶333-36.)

C. Claim 6

DTB discloses claim 6. (*Supra* §VI.F; EX-1002 ¶337.) As in DTB, McCartney's bookmark could be based on the text stream *or* the audio stream. (EX-1002 ¶337.)

A POSITA would have been motivated to combine McCartney and DTB and would have had a reasonable expectation of success in doing so. (*Supra* §§XV.A.1-XV.A.2; EX-1002 ¶338**.)**

Thus, McCartney and DTB render claim 6 obvious. (*Supra* §XIII; EX-1002 ¶¶337-39.)

D. Claim 13

DTB discloses the limitations of claim 13. (*Supra* §VI.M; EX-1002 ¶340.)

A POSITA would have understood that the DTB/McCartney bookmark would include either the ID of the TSD element or the absolute time being marked as well as an identification of the TSD file. (*Supra* §XV.A.2; EX-1002 ¶341.) Both options determine the position using both the bookmark (for the position) and the TSD time offsets (as determined by either ID or relative time) to identify the appropriate audio clip and text associated with that audio clip. (EX-1002 ¶341.)

A POSITA would have been motivated to combine McCartney and DTB and would have had a reasonable expectation of success in doing so. (*Supra* §§XV.A-XV.B, XIII.A.11 (TSD file stores synchronization time offsets); EX-1002 ¶342.)

Thus, McCartney and DTB render claim 13 obvious. (*Supra* §XIII; EX-1002 ¶¶340-43.)

XVI. GROUND 2D: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND DUNCAN.

McCartney discloses visually rendered content (e.g., text). (EX-1013 ¶¶[0044]-[0046]; EX-1002 ¶345; *supra* §XIIIA.) The remaining limitation of claim 7 is disclosed by Duncan. (*Supra* §VIII; EX-1002 ¶346.) It would have been obvious to a POSITA to use Duncan's auto page turn feature on McCartney's review at the TOAC computer. (*Id.*)

A POSITA would have been motivated to automatically turn pages in McCartney's system. First, McCartney discloses a method for producing digital talking books, and Duncan discloses a method for presenting them. (EX-1013, Abstract; EX-1027, Abstract; EX-1002 ¶348.) Second, use of Duncan's system with the method of McCartney would allow McCartney's TOAC operator or subsequent DTB player to render audio and text "in the familiar page-by-page manner," which is a "desirable" capability for electronic book readers and would enable efficient verification of audio and text synchronization. (EX-1027 ¶[0042]; EX-1013 ¶[0009] (TOAC operator can verify that text and audio have been correctly linked); EX-1002 ¶349.)

Third, the combination represents merely the simple addition of one known element (e.g., automatic page turns) to another known element (e.g., McCartney's rendering) to obtain predictable results (e.g., rendering content using automatic page turns). (EX-1002 ¶350); *KSR*, 550 U.S. at 417.

Fourth, the combination uses a known technique (e.g., providing automated page turns) to improve a similar method (e.g., McCartney's rendering of book content) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., automatic page turning) to a known method (e.g., rendering content) that is ready for improvement and yields predictable results (e.g., rendering content with automatic page turning). (*Id.*)

A POSITA would have had a reasonable expectation of success in making the modification at least because Duncan's automatic page turn feature advances the viewer to the next page in an XML text file, (EX-1027 ¶¶[0041]-[0042]), and McCartney discloses identifying page events in an XML text stream. (EX-1013 ¶[0041] (method of McCartney correlates "page elements" in the text with "page events" in the audio); *id.* ¶[0029] (text can be in XML format); EX-1002 ¶351.)

Thus, McCartney and Duncan render claim 7 obvious. (*Supra* §§XIII.A, XIII.C; EX-1002 ¶¶344-52.)

XVII. GROUND 2E: CLAIM 8 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND WOODILL.

Even if Woodill's disclosure were not incorporated in McCartney's disclosure, a POSITA would have been motivated to combine McCartney and Woodill for several reasons. First, McCartney explicitly recognizes that its TSD file, which includes the plurality of synchronization time offsets, can be generated by Woodill's method. (EX-1013 ¶[0022]; EX-1002 ¶356.) Second, both references are directed to creating TSD files. (EX-1014 ¶[0024]; EX-1013, Abstract; EX-1002 ¶356.) Third, McCartney and Woodill share inventors. (EX-1013, 1; EX-1014, 1.) A POSITA would have had a reasonable expectation of success in combining McCartney and Woodill for the same reasons. (EX-1002 ¶356.)

Thus, McCartney and Woodill render claim 8 obvious. (*Supra* §XIII.D; *Id.*

¶¶353-57.)

XVIII. GROUND 2F: CLAIMS 14 AND 20 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND BULTERMAN.

McCartney discloses or renders obvious claim 14. (*Supra* §XIII.) Claims 14 and 20 were also obvious in further view of Bulterman. (EX-1002 ¶¶358-64.)

A. Claim 14

A POSITA would have been motivated to include the actual text (as opposed to a link, path, or ID) in McCartney's TSD and/or BPM file. (*Supra* §X.A; EX-1101, 5; EX-1002 ¶360.) A POSITA would have had a reasonable expectation of success in applying Bulterman's method to a TSD and/or BPM file because it represents mere modification of an attribute already present in those files. (EX-1101, 2-5; EX-1013, 7, 12; EX-1002 ¶360.) Additionally, the combination applies a known technique (Bulterman's inclusion of text in a "source" field of a synchronization file) to a known element (McCartney's synchronization files) that is ready for improvement and yields a predictable result (a synchronization file including human readable text). Further, the compatibility of information contained in XML files (e.g., TSD and BPM files) with SMIL files was known in the art. (EX-1002 ¶360 (citing EX-1013 ¶[0057]; *id.* ¶[0030], 7; EX-1003, 23 (DTB SMIL files are "XML documents")).)

Thus, McCartney and Bulterman render claim 14 obvious. (EX-1002 ¶¶359-61.)

B. Claim 20

Including the actual text in McCartney's BPM "path" field (e.g., McCartney's location information) would have been obvious to a POSITA. (*Supra* §X.A; EX-1002 ¶363.) The presence of text (as opposed to a file path) would indicate to a POSITA that the text stream is located in the BPM file. (EX-1002 ¶363.) Alternatively, it would have been obvious to a POSITA to include an indicator like that disclosed in Bulterman. (*Id.*) A POSITA would have been motivated to combine Bulterman's indicator with McCartney's path information, and would have had a reasonable expectation of success in doing so. (*Supra* §§X.A, XVIII.A; EX-1002 ¶363.)

Thus, McCartney and Bulterman render claim 20 obvious. (EX-1002 ¶¶362-64.)

XIX. GROUND 2G: CLAIM 15 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY AND SHTEYN.

Shteyn discloses that content can be split into parts and delivery of those parts can be coordinated from a server to a client device in dependence on the semantics of the content, for example, "the end of musical phrase, paragraph, target control device, etc." (EX-1086, Abstract, 3:1-2, 3:27-4:25, claim 1; EX-1002 ¶366.)

It would have been obvious to a POSITA to "split" McCartney's audio files "into parts" and coordinate the delivery of those parts consistent with synchronization time offsets according to the method of Shteyn. (*Id.*; EX-1002 ¶367.) The benefits of doing so are explained above. (*Supra* §XI; EX-1002 ¶367.) Moreover, Shteyn explains benefits including "low or negligible play-out latency[,]" "relatively short download time" of each segment, "simultaneous play-out and downloading do not significantly compete for the same system resources[,]" "[t]he parts can have different data formats[,]" "[c]ontent parts can physically reside on different servers[,]" etc. (EX-1086, 2:11-3:7; EX-1002 ¶367.) A POSITA would have had a reasonable expectation of modifying McCartney according to the method of Shteyn because division and data and processing of split data was well-known in the art. (*See, e.g.*, EX-1086, Abstract, 1:13-14, 2:20-22 (explaining that "splitters" and clients for processing split data were available); EX-1002 ¶367.) Thus, McCartney and Shteyn render claim 15 obvious. (*Supra* §XIII.A; EX-1002 ¶¶365-68.)

XX. GROUND 2H: CLAIM 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF MCCARTNEY, SHTEYN, AND COPLEY.

Copley discloses or renders obvious the limitation of claim 16. (*Supra* §XII.)

It would have been obvious to a POSITA to manage retrieval of the data in McCartney by Copley's method. (*Id.*; EX-1002 ¶370.) In addition to the reasons discussed above, a POSITA would have been motivated to combine McCartney and

Copley with a reasonable expectation of success because McCartney discloses consumption of audio and text media, and Copley provides a method of doing so with "a high quality of service." (EX-1013 ¶[0046]; EX-1009 ¶¶[0013]-[0015]; EX-1002 ¶370.) Thus, McCartney, Shteyn, and Copley render claim 16 obvious. (*Supra* §§XIII.A, XIX; EX-1002 ¶¶369-71.)

XXI. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

Where, as here, a strong *prima facie* obviousness showing exists, secondary considerations may not dislodge the obviousness conclusion. *Leapfrog Enters., Inc. v. Fisher-Price, Inc.,* 485 F.3d 1157, 1162 (Fed. Cir. 2007). Petitioners are aware of no evidence supporting a claim for secondary considerations.

XXII. CONCLUSION

Amazon requests the Board institute trial and cancel all challenged claims.[3]

XXIII. MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT

Pursuant to 37 C.F.R. §42.8(a)(1), the mandatory notices identified in 37 C.F.R. §42.8(b) are provided below as part of this Petition.

[3] Petitioners will address discretionary denial issues if raised by PO. *See* Memorandum from Acting Director Stewart, *Interim Processes for PTAB Workload Management* (March 26, 2025).

A. Real Parties-In-Interest (37 C.F.R. §42.8(b)(1))

Amazon.com, Inc., Amazon.com Services LLC, Amazon Web Services, Inc., and Audible, Inc. are the real parties-in-interest.

B. Related Matters (37 C.F.R. §42.8(b)(2))

PO asserted the '907 patent against Petitioners in a patent infringement lawsuit captioned *Audio Pod IP, LLC v. Amazon.com, Inc. et al.*, No. 3-24-cv-00407 (E.D. Va., filed May 30, 2024)[4].

If this IPR is instituted and the above proceeding is not stayed, Petitioners hereby stipulate not to pursue in that proceeding any ground of invalidity, against any claim challenged herein, that was raised or reasonably could have been raised in this Petition.

Petitioners further filed the following IPR petitions challenging claims of related patents:

Case Number	Challenged Patent	Challenged Claims
IPR2025-00757	U.S. Pat. No. 10,091,266	1-13
IPR2025-00765	U.S. Pat. No. 8,738,740	12-17
IPR2025-00768	U.S. Pat. No. 10,805,111	1-16
IPR2025-00769	U.S. Pat. No. 9,954,922	1-20
IPR2025-00774	U.S. Pat. No. 8,738,740	1-6, 10-11, 18
IPR2025-00777	U.S. Pat. No. 9.319.720	1-29

[4] *Audio Pod IP, LLC v. Amazon.com, Inc. et al.*, No. 1-24-cv-00444 (E.D. Va., filed March 20, 2024) was consolidated with this case.

C. Lead and Backup Counsel (37 C.F.R. §42.8(b)(3))

Petitioners provide the following designation of counsel, all of whom are included in Customer No. 20,995 identified in Petitioners' Power of Attorney.

Lead Counsel	**Back-up Counsel**
Colin B. Heideman (Reg. No. 61,513) 2cbh@knobbe.com BoxSEAZNL2185L2LP@knobbe.com Postal and Hand-Delivery Address: Knobbe, Martens, Olson, & Bear, LLP 555 110th Ave. NE, Ste. 500 Bellevue, WA 98004 Telephone: (206) 405-2000 Facsimile: (206) 405-2001	Joseph R. Re (Reg. No. 31,291) 2jrr@knobbe.com Postal and Hand-Delivery Address: Knobbe, Martens, Olson, & Bear, LLP 2040 Main Street, 14th Floor Irvine, CA 92614 Telephone: (949) 760-0404 Facsimile: (949) 760-9502 Christie R.W. Matthaei (Reg. No. 62,933) 2crw@knobbe.com Nathan D. Reeves (Reg. No. 77,806) 2ndr@knobbe.com Postal and Hand-Delivery Address: Knobbe, Martens, Olson & Bear, LLP 555 110th Ave. NE, Ste. 500 Bellevue, WA 98004 Telephone: (206) 405-2000 Facsimile: (206) 405-2001 Daniel Hughes (Reg. No. 76,592) 2dph@knobbe.com Postal and Hand-Delivery Address: Knobbe, Martens, Olson & Bear, LLP 3579 Valley Centre Dr., Ste. 300 San Diego, CA 92130 Telephone: (858) 707-4000 Facsimile: (858) 707-4001

D. Service Information (37 C.F.R. §42.8(b)(4))

Please direct all correspondence to lead counsel and back-up counsel at the addresses shown above. Petitioners also consent to electronic service by email to BoxSEAZNL2185L2LP@knobbe.com.

E. Grounds for Standing (37 C.F.R. §42.104(a))

Petitioners certify that the '907 patent is available for IPR and that Petitioners are not barred or estopped from requesting IPR on the identified grounds. This petition is being filed within one year of service of the original complaint against Petitioners in the district court litigation.

F. Payment of Fees (37 C.F.R. §42.103)

The Office may charge the §42.15(a) fee to Deposit Account No. 11-1410. Review of twenty-one claims is requested. Payment for any additional fees due may be charged to the above-referenced Deposit Account.

Respectfully submitted,

KNOBBE MARTENS OLSON & BEAR, LLP

Dated: May 13, 2025

/Colin B. Heideman /

Colin B. Heideman (Reg. No. 61,513)
Joseph R. Re (Reg. No. 31,291)
Christie R.W. Matthaei (Reg. No. 62,933)
Nathan D. Reeves (Reg. No. 77,806)
Daniel Hughes (Reg. No. 76,592)

Counsel for Petitioners
Amazon.com, Inc.,
Amazon.com Services LLC,
Amazon Web Services, Inc., and
Audible, Inc.

APPENDIX

Listing of Claims from U.S. 9,729,907	
Claim 1	
1[pre]	A method comprising:
1[a]	creating a descriptor file for synchronizing a plurality of digital media streams, wherein the plurality of digital media streams each contain digital media content corresponding to a same originating work, wherein the plurality of digital media streams includes a first digital media stream containing a digital audio narration of the originating work and one or more other digital media streams, and wherein the descriptor file is external to the first digital media stream;
1[b]	storing location information for the plurality of digital media streams in the descriptor file;
1[c]	identifying a plurality of time offsets in a timeline of the digital audio narration of the first digital media stream, wherein the plurality of time offsets correspond to a plurality of content points in the digital audio narration;
1[d]	storing the plurality of time offsets and the plurality of content points in the descriptor file in a manner indicating a correlation between the plurality of time offsets and the plurality of content points;
1[e]	identifying synchronization points in the digital media content of the one or more other digital media streams;
1[f]	selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets; and
1[g]	storing the synchronization time offsets and the synchronization points in the descriptor file in a manner indicating a correlation between the synchronization time offsets and the synchronization points, such that the descriptor file allows a synchronized rendering of the plurality of digital media streams on a client device.

Listing of Claims from U.S. 9,729,907	
Claim 2	
--	The method of claim 1, further comprising: storing the descriptor file on a server that is accessible to the client device.
Claim 3	
--	The method of claim 1, further comprising: simultaneously rendering two or more of the plurality of digital media streams on the client device by using the descriptor file to synchronize the two or more of the plurality of digital media streams.
Claim 4	
4[a]	The method of claim 3, further comprising: tracking a current position in at least one of the two or more of the plurality of digital media streams as the two or more of the plurality of digital media streams are rendered; and
4[b]	creating a bookmark by setting the current position as a bookmarked position, wherein the bookmark includes the bookmarked position and identifies the descriptor file.
Claim 5	
--	The method of claim 4, further comprising: rendering one or more of the plurality of digital media streams starting from the bookmarked position on the client device or on a second other client device.
Claim 6	
--	The method of claim 5, wherein the one or more of the plurality of digital media streams were not used to create the bookmark.
Claim 7	
7[a]	The method of claim 3, wherein at least one of the two or more of the plurality of digital media streams contains visually rendered digital media content, the method further comprising:

Listing of Claims from U.S. 9,729,907	
7[b]	providing automated page turns as the two or more of the plurality of digital media streams are simultaneously rendered.
Claim 8	
--	The method of claim 1, wherein at least some of the synchronization time offsets are selected in dependence on natural language gaps, inter-word gaps, punctuation marks, or parts of speech.
Claim 9	
--	The method of claim 1, wherein the first digital media stream includes a plurality of content segments, and wherein the plurality of content segments are defined by the plurality of time offsets in the descriptor file.
Claim 10	
--	The method of claim 9, wherein the plurality of time offsets includes start times of the content segments, end times of the content segments, durations of the content segments, or a combination thereof.
Claim 11	
11[a]	The method of claim 1, further comprising: selecting two or more of the plurality of digital media streams;
11[b]	determining a first position in the digital media content of the two or more of the plurality of digital media streams in dependence on the synchronization time offsets in the descriptor file; and
11[c]	simultaneously rendering the two or more of the plurality of digital media streams on the client device starting from the first position by using the descriptor file to synchronize the two or more of the plurality of digital media streams.

Listing of Claims from U.S. 9,729,907	
Claim 12	
--	The method of claim 11, wherein determining the first position comprises detecting a user interaction with a media rendering utility or a user interface on the client device.
Claim 13	
--	The method of claim 11, wherein determining the first position comprises determining the first position in dependence on the synchronization time offsets in the descriptor file and in dependence on a bookmarked position or a default position.
Claim 14	
--	The method of claim 1, wherein the descriptor file contains information that is human readable when rendered.
Claim 15	
--	The method of claim 1, further comprising: coordinating delivery of the digital media content of the plurality of digital media streams from one or more servers to the client device in dependence on the synchronization time offsets in the descriptor file.
Claim 16	
--	The method of claim 15, further comprising: managing network connections between the client device and the one or more servers in dependence on service failure, service degradation, digital data throughput rate, or a combination thereof.
Claim 17	
--	The method of claim 1, wherein the digital media content of each of the plurality of digital media streams includes an audio recording, a video recording, an audio/video recording, a static image, a moving image, an animation, an illustration, text content, footnote content, quotation content, user-entered content, advertising content, or a combination thereof.

Listing of Claims from U.S. 9,729,907	
Claim 18	
--	The method of claim 1, further comprising: navigating one or more of the plurality of digital media streams by using the descriptor file on the client device, wherein at least some of the digital media content of the one or more of the plurality of digital media streams is not resident on the client device.
Claim 19	
	The method of claim 1, wherein one or more of the plurality of digital media streams are discontinuous media streams that are discontinuous when rendered.
Claim 20	
--	The method of claim 1, wherein the location information for one or more of the plurality of digital media streams indicates that the one or more of the plurality of digital media streams are located within the descriptor file.
Claim 21	
21[pre]	A non-transitory computer-readable storage medium having stored thereon instructions which, when executed by a processor of a computing device, cause the computing device to:
21[a]	create a descriptor file for synchronizing a plurality of digital media streams, wherein the plurality of digital media streams each contain digital media content corresponding to a same originating work, wherein the plurality of digital media streams includes a first digital media stream containing a digital audio narration of the originating work and one or more other digital media streams, and wherein the descriptor file is external to the first digital media stream;
21[b]	store location information for the plurality of digital media streams in the descriptor file;

Listing of Claims from U.S. 9,729,907	
21[c]	identify a plurality of time offsets in a timeline of the digital audio narration of the first digital media stream, wherein the plurality of time offsets correspond to a plurality of content points in the digital audio narration;
21[d]	store the plurality of time offsets and the plurality of content points in the descriptor file in a manner indicating a correlation between the plurality of time offsets and the plurality of content points;
21[e]	identify synchronization points in the digital media content of the one or more other digital media streams;
21[f]	select synchronization time offsets that correspond to the synchronization points from the plurality of time offsets; and
21[g]	store the synchronization time offsets and the synchronization points in the descriptor file in a manner indicating a correlation between the synchronization time offsets and the synchronization points, such that the descriptor file allows a synchronized rendering of the plurality of digital media streams on a client device.

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this **PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 9,729,907** contains 13,883 words according to the word-processing program used to prepare this paper. The foregoing word count complies with the 14,000-word type-volume limit specified by 37 C.F.R. §42.24(a)(1).

Dated: May 13, 2025

By: */Colin B. Heideman /*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on the date below a copy of this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,729,907** and **ACCOMPANYING EXHIBITS** are being served on May 13, 2025 via Federal Express overnight mail on counsel of record for U.S. Patent No. 9,729,907 at the Correspondence Address of record below:

Neil Teitelbaum
Teitelbaum & Bouevitch
834 Colonel By Drive
Ottawa, ON K1S 5C4
Canada

A courtesy copy is also being served via email on counsel for the patent holder in the pending district court litigation:

Daignault Iyer LLP
8229 Boone Blvd., Suite 450
Vienna, VA 22182
Email : DI-AudioPod@daignaultiyer.com

Dated: May 13, 2025 By: */Colin B. Heideman /*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP