

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

DECLARATION OF PROFESSOR KETAN MAYER-PATEL, Ph.D.

TABLE OF CONTENTS

I.	BACKGROUND -----	1
A.	Experience and Qualifications-----	1
B.	Materials Considered -----	3
II.	APPLICABLE LEGAL STANDARDS -----	6
A.	Claim Construction -----	7
B.	Obviousness -----	8
III.	PERSON OF ORDINARY SKILL IN THE ART -----	12
IV.	TECHNOLOGY BACKGROUND -----	14
V.	THE '907 PATENT -----	15
A.	Overview -----	15
B.	Claims -----	16
VI.	CLAIMS 1-21 OF THE '907 PATENT WOULD HAVE BEEN OBVIOUS -----	18
A.	Claims 1-14, 17-19, and 21 Would Have Been Obvious in View of DTB -----	18
1.	Claim 1 -----	18
a.	Preamble -----	18
b.	Element 1[a][i]: Descriptor File -----	18
c.	Elements 1[a][ii]-1[a][iii]: Digital Media Streams Corresponding to a Same Originating Work-----	19
d.	Element 1[a][iv]: Descriptor File External to the Audio Stream -----	20

e.	Element 1[b]: Location Information -----	21
f.	Element 1[c][i]: Time Offsets-----	23
g.	Element 1[c][ii]: Content Points in Audio Narration-----	25
h.	Element 1[d]: Correlating Time Offsets and Content Points-----	26
i.	Element 1[e]: Synchronization Points -----	27
j.	Element 1[f]: Synchronization Time Offsets-----	30
k.	Element 1[g][i]: Correlating Synchronization Time Offsets and Synchronization Points -----	31
l.	Element 1[g][ii]: Synchronized Rendering-----	32
2.	Claim 2 -----	32
3.	Claim 3 -----	34
4.	Claim 4 -----	34
5.	Claim 5 -----	37
6.	Claim 6 -----	38
7.	Claim 7 -----	39
a.	Element 7[a]: Visually Rendered Content-----	39
b.	Element 7[b]: Automated Page Turns -----	39
8.	Claim 8 -----	40
9.	Claim 9 -----	40
10.	Claim 10-----	42
11.	Claim 11 -----	42
a.	Element 11[a]: Selecting Two or More of the Streams-----	42

b.	Element 11[b]: Determining a First Position-----	43
c.	Element 11[c]: Rendering from the First Position-----	44
12.	Claim 12-----	45
13.	Claim 13-----	46
14.	Claim 14-----	46
15.	Claim 17-----	47
16.	Claim 18-----	48
a.	Element 18[a][i]: Navigating-----	48
b.	Element 18[a][ii]: Some Content Not Resident on Client Device -----	49
17.	Claim 19-----	49
18.	Claim 21-----	50
B.	Claims 2 and 18-19 Would Have Been Obvious in View of DTB and Yoshimura-----	53
1.	Claim 2-----	53
2.	Claim 18-----	55
a.	Element 18[a][i]: Navigating-----	55
b.	Element 18[a][ii]: Some Content Not Resident on Client Device -----	57
3.	Claim 19-----	58

C.	Claim 7 Would Have Been Obvious in View of DTB and Duncan -----	59
D.	Claim 8 Would Have Been Obvious in View of DTB and Heckerman -----	61
E.	Claims 14 and 20 Would Have Been Obvious in View of DTB and Bulterman -----	64
1.	Claim 14 -----	65
2.	Claim 20 -----	68
F.	Claim 15 Would Have Been Obvious in View of DTB and Yang -----	68
G.	Claim 16 Would Have Been Obvious in View of DTB, Yang, and Copley -----	70
H.	Claims 1-3, 8-12, 14, 17-19, and 21 Would Have Been Obvious in View of McCartney -----	73
1.	Claim 1 -----	73
a.	Preamble -----	73
b.	Element 1[a][i]: Descriptor File -----	73
c.	Element 1[a][ii]: Same Originating Work -----	75
d.	Element 1[a][iii]: Audio Stream and Other Digital Media Stream -----	76
e.	Element 1[a][iv]: Descriptor File External to the Audio Stream -----	76
f.	Element 1[b]: Location Information -----	77
i.	TSD File -----	77
ii.	BPM File -----	78
g.	Element 1[c][i]: Time Offsets -----	79

h.	Elements 1[c][ii] and 1[d]: Content Points in Audio Narration Correlated to Time Offsets-----	80
i.	Element 1[e]: Synchronization Points -----	82
j.	Element 1[f]: Synchronization Time Offsets-----	83
k.	Element 1[g][i]: Correlating Synchronization Time Offsets and Synchronization Points -----	85
l.	Element 1[g][ii]: Synchronized Rendering-----	86
2.	Claim 2 -----	87
3.	Claim 3 -----	88
4.	Claim 8 -----	89
5.	Claims 9-10 -----	90
6.	Claims 11-12-----	91
a.	Element 11[a]: Selecting Two or More of the Streams-----	91
b.	Element 11[b]: Determining a First Position and Element 12: User Interaction-----	92
c.	Element 11[c]: Rendering from the First Position-----	93
7.	Claim 14-----	93
8.	Claim 17-----	94
9.	Claim 18-----	95
10.	Claim 19-----	96
11.	Claim 21 -----	97
I.	Claims 2 and 18 Would Have Been Obvious in View of McCartney and Kauffman -----	97
1.	Claim 2 -----	98

2.	Claim 18 -----	99
J.	Claims 4-6 and 13 Would Have Been Obvious in View of McCartney and DTB -----	100
1.	Claim 4 -----	101
a.	Element 4[a]: Tracking Current Position -----	101
b.	Element 4[b]: Bookmark -----	102
2.	Claim 5 -----	103
3.	Claim 6 -----	104
4.	Claim 13 -----	105
K.	Claim 7 Would Have Been Obvious in View of McCartney and Duncan -----	106
L.	Claim 8 Would Have Been Obvious in View of McCartney and Woodill -----	108
M.	Claims 14 and 20 Would Have Been Obvious in View of McCartney and Bulterman -----	109
1.	Claim 14 -----	109
2.	Claim 20 -----	110
N.	Claim 15 Would Have Been Obvious in View of McCartney and Shteyn -----	111
O.	Claim 16 Would Have Been Obvious in View of McCartney, Shteyn, and Copley -----	113
VII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	114
VIII.	CONCLUSION -----	114

I, Ketan Mayer-Patel, do hereby declare:

1. I am making this declaration at the request of Petitioners Amazon.com, Inc., Amazon.com Services LLC, Amazon Web Services, Inc., and Audible, Inc. (“Petitioners”). I have been retained by Petitioners as a technical expert in this matter.

2. I am being compensated for my work on this case. My compensation does not depend on the content of this Declaration or the outcome of these proceedings.

I. BACKGROUND

A. Experience and Qualifications

3. I received Bachelor of Arts degrees in Computer Science and Economics in 1992, a Master of Science in 1997 from the Department of Electrical Engineering and Computer Science, and a Ph.D. in 1999 from the Department of Electrical Engineering and Computer Science, all from the University of California, Berkeley.

4. I have been involved in the research and development of multimedia computing systems for nearly 30 years. I have been a faculty member at the University of North Carolina since January 2000, where I perform research and teach in the areas of networking, web programming, and multimedia computing. I also have

expertise other areas, including distributed systems, networking devices, and the general operation of computer systems.

5. I am a member of the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE). These are the two leading professional societies for both academic and practicing computer scientists.

6. I have authored or co-authored over 30 papers in peer-reviewed journals and conference proceedings. I have served as an Associate Editor for both IEEE Transactions on Multimedia and ACM Transactions on Multimedia Computing, Communications, and Applications, which are the two leading journals in the field. I regularly serve as a member of the technical program committee for a number of different conferences and workshops including ACM Multimedia, The International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), IFIP Networking, ACM Multimedia Systems (MMSys), MMEDIA, and SIGMAP. I am also currently chair of the standing executive committee for both NOSSDAV and MMSys. A complete listing of all my publications can be found in my CV, which I understand is being submitted as Exhibit 1096. I am also a named inventor or co-inventor on multiple issued patents, which are also listed in my CV.

7. My research has been supported by both government agencies as well as private industry. I received the National Science Foundation (NSF) CAREER Award in 2003 while an Assistant Professor. I have been a principal investigator for

grants awarded by the NSF, the Office of Naval Research, and the Laboratory of Analytic Sciences. I have also served on several NSF reviewing panels for funding recommendations.

8. In my research and teaching I have considered problems of video streaming, dynamic adaptation and transcoding of media, adaptive streaming transport protocols, telepresence, and scalable display architectures, among others.

9. In the classroom, I have regularly taught classes on Data Structures, Foundations of Programming, Modern Web Programming, Files and Databases, and Multimedia Computing and Networking. I also serve as the Director of Undergraduate Studies for the Department of Computer Science.

B. Materials Considered

10. In preparing this Declaration, I have considered the following materials:

Exhibit No.	Description
1001	U.S. Patent No. 9,729,907 (“the ’907 patent”)
1003	NATIONAL INFORMATION STANDARDS ORGANIZATION, SPECIFICATIONS FOR THE DIGITAL TALKING BOOK (ANSI/NISO Z39.86-2002) (2002) (“DTB”)
1006	Yoshimura et al., <i>Content Delivery Network Architecture for Mobile Streaming Service Enabled by SMIL Modification</i> , 86 IEICE TRANSACTIONS ON COMM’N 1778 (2003) (“Yoshimura”)

Amazon.com, Inc. v. Audio Pod IP, LLC
Declaration of Ketan Mayer-Patel – U.S. Pat. No. 9,729,907

Exhibit No.	Description
1007	Excerpts from DICK C.A. BULTERMAN & LLOYD RUTLEDGE, SMIL 2.0, INTERACTIVE MULTIMEDIA FOR WEB AND MOBILE DEVICES (2004) (“Bulterman”)
1009	U.S. Patent Publication No. 2003/0061305 (“Copley”)
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”)
1027	U.S. Patent Publication No. 2003/0013073 (“Duncan”)
1029	Robert Steele et al., <i>Accessing SMIL-based Dynamically Adaptable Multimedia Presentations from Mobile Devices</i> , in INT’L CONF. ON INFO. TECH.: CODING AND COMPUTING (2004) (“Steele”)
1030	U.S. Patent No. 6,260,011 (“Heckerman”)
1032	U.S. Patent. No. 6,108,703 (“Leighton”)
1033	U.S. Patent No. 5,922,045 (“Hanson”)
1045	Chun-Chuan Yang et al., <i>Design and Implementation of the Just-in-Time Retrieving Policy for Schedule-Based Distributed Multimedia Presentations</i> , 71 J. SYS. & SOFTWARE 49 (2004) (“Yang”)
1046	Tredinnick, <i>Implementing the SMIL Specification</i> , PROC. OF THE LINUX SYMP. (2003) (“Tredinnick”)
1049	UNIFORM RESOURCE IDENTIFIERS (URI): GENERIC SYNTAX, RFC 2396 (1998)
1051	Excerpts from World Wide Web Consortium, <i>Synchronized Multimedia Integration Language (SMIL 2.0) Specification</i> , W3C Recommendation (Aug. 7, 2001),

Amazon.com, Inc. v. Audio Pod IP, LLC
Declaration of Ketan Mayer-Patel – U.S. Pat. No. 9,729,907

Exhibit No.	Description
	https://www.w3.org/TR/2001/REC-smil20-20010807/smil20.html . (the “SMIL Standard”)
1057	U.S. Patent No. 5,798,841 (“Takahashi”)
1058	U.S. Patent No. 5,021,989 (“Fujisawa”)
1059	PCT Patent Publication No. WO2001/01373 (“Hendricks”)
1063	P. Delacourt & C.J. Wellekens, <i>DISTBIC: A speaker-based segmentation for audio data indexing</i> , 32 SPEECH COMM’N 111 (2000) (“Delacourt”)
1069	PCT Patent Publication No. WO2002/080524 (“Dimitrova”)
1072	Excerpt from File History for European Patent Publication No. EP1961154
1074	Final Office Action dated July 28, 2010, U.S. Patent App. No. 12/096933
1075	Barry Arons, <i>SpeechSkimmer: A System for Interactively Skimming Recorded Speech</i> , 4 ACM TRANSACTIONS ON COMPUTER-HUMAN INTERACTION 3 (1997) (“Arons”)
1086	PCT Patent Publication No. WO2001/024474 (“Shteyn”)
1095	Excerpts from the File History for U.S. Patent No. 9,729,907
1101	John E. Simpson, <i>From XML to SMIL</i> , 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., <i>Streaming Technology in 3G Mobile Communication Systems</i> , 34 COMPUTER, no. 9, at 46-52 (2001) (“Elsen”)
1103	WIKIBOOKS, <i>XML – Managing Data Exchange/SMIL</i> , https://en.wikibooks.org/w/index.php?title=XML -

Exhibit No.	Description
	<u>Managing Data Exchange/SMIL&oldid=283559</u> (November 15, 2005 snapshot via Wikibooks) (“SMIL Wiki”)
1104	Lloyd Rutledge, <i>SMIL 2.0: XML for Web Multimedia</i> , IEEE INTERNET COMPUTING, Sept.-Oct. 2001, at 78-84 (“Rutledge”)
1105	Dongkyoo Shin et al., <i>Design and Implementation of the SMIL (Synchronized Multimedia Integration Language) Player</i> , 48 IEEE TRANSACTIONS ON CONSUMER ELECS. 575 (2002) (“Shin”)
1106	Herng-Yow Chen et al., <i>Design of a Web-Based Synchronized Multimedia Lecture System for Distance Education</i> , 2 PROC. IEEE INT’L CONF. ON MULTIMEDIA COMPUTING AND SYS. 887 (1999) (“Chen”)
1107	U.S. Patent Publication No. 2007/0124331 (“Griffin”)

11. In addition, I have reviewed the full file history of the ’907 patent. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the ’907 patent.

II. APPLICABLE LEGAL STANDARDS

12. I have been asked to provide my opinion as to whether the claims of the ’907 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention, in view of the prior art.

13. I am a computer scientist by training and profession. The opinions I am expressing in this report involve the application of my training and technical

knowledge and experience to the evaluation of certain prior art with respect to the '907 patent.

14. Although I have been involved as a technical expert in patent matters before, I am not an expert in patent law. Therefore, the attorneys from Knobbe, Martens, Olson & Bear, LLP have provided me with guidance as to the applicable patent law in this matter. The paragraphs below express my understanding of how I must apply current principles related to patent validity to my analysis.

A. Claim Construction

15. It is my understanding that in determining whether a patent claim is obvious in view of the prior art, the Patent Office construes the claim by giving the claim terms their plain and ordinary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention in view of the intrinsic record (patent specification and file history). For the purposes of this review, and to the extent necessary, I have interpreted each claim term in accordance with its plain and ordinary meaning as it would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record. I understand that the time of the invention is December 13, 2005.

16. I understand that a patent and its prosecution history are considered “intrinsic evidence” and are the most important sources for interpreting claim language in a patent. I also understand that in reading the claim, I must not import limitations

from the specification into the claim terms; in other words, I must not narrow the scope of the claim terms by implicitly adding disclosed limitations that have no express basis in the claims. The prosecution history of related patents and applications can also be relevant.

17. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

18. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of ordinary skill in the art would have given to terms in December 2005 based on a review of the intrinsic evidence.

B. Obviousness

19. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the ’907 patent and

the prior art; (iii) the level of ordinary skill in the art in December 2005; and (iv) additional considerations, if any, that indicate that the invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” of non-obviousness or obviousness.

20. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would have known in December 2005. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

21. It is my understanding that something is “inherent in,” and therefore taught by, the prior art, if it necessarily flows from the explicit disclosure of the prior art. I understand that the fact that a certain result or characteristic *may be* present in the prior art is not sufficient to establish inherency. However, if the result or characteristic is *necessarily* present based upon the explicit disclosure in the prior art, it is inherent in the prior art and is therefore disclosed.

22. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor’s attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an

obviousness analysis. Here, all of the references relied on in my obviousness analysis below are from the same field of endeavor as the '907 patent, e.g., content distribution and/or rendering. The references are also pertinent to a particular problem the inventor was focused on, e.g., efficient and effective distribution and/or rendering of content.

23. It is my understanding that the law recognizes several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- a predictable use of prior art elements according to their established functions;
- using known techniques to improve similar devices (methods, or products) in the same way;
- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;

- choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success (in which case a claim would have been obvious to try);
- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

24. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts; (4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was

within the knowledge of a person of ordinary skill as of December 13, 2005. I am not aware of any evidence of secondary considerations that would suggest that the claims of the '907 patent would have been nonobvious in December 2005.

25. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered from the perspective of a person of ordinary skill at the time the alleged invention was made – December 2005 in this case.

26. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

III. PERSON OF ORDINARY SKILL IN THE ART

27. It is my understanding that when interpreting the claims of the '907 patent and evaluating whether a claim would have been obvious, I must do so based on the perspective of a person of ordinary skill in the art at the relevant priority date. I understand that the relevant priority date of the '907 patent is December 13, 2005.

28. I understand that in determining the level of ordinary skill in the art, several factors are considered. Those factors may include: (i) the type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. A person of ordinary skill in the art must have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

29. The '907 patent describes the use of well-known technologies for the rendering and/or distribution of digital content. Based on my review of the specification and claims of the '907 patent, it is my opinion that a person of ordinary skill in the art would have had a minimum of 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. Work experience could substitute for formal education and additional formal education could substitute for work experience.

30. My conclusions below that the claims of the '907 patent would have been obvious would remain the same even if the priority date, field of endeavor, or level of ordinary skill were slightly different.

31. I meet the above definition of a person of ordinary skill in the art, and did so as of December 13, 2005. Also, I have worked with persons of ordinary skill

in the art through my professional and academic experiences, and I have an understanding of their skill level around December 2005.

IV. TECHNOLOGY BACKGROUND

32. The '907 patent's 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 December 2005. By 2005, synchronization files were so common there were multiple industry standards implementing synchronization files to ensure interoperability and compatibility between multimedia products and services looking to synchronize multiple media streams. 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 (Tredinnick), 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 (DTB), vii, 6-7.)

33. Other synchronization files for "talking books"—now commonly referred to as "audiobooks"—were also well known before 2005. 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 various “synchronization files.” (EX-1013 (McCartney), Abstract.)

V. THE '907 PATENT

A. Overview

34. 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:

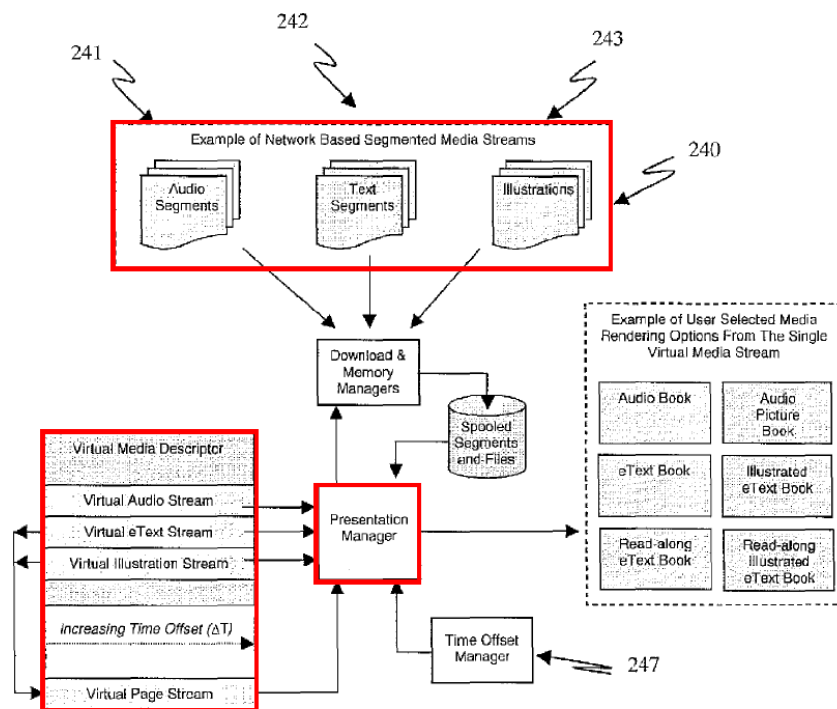


Fig. 24a

(EX-1001 ('907 patent), Fig. 24a¹.)

¹ Images in this Declaration may be colored and/or annotated for clarity.

35. The Virtual Media Descriptor includes a “Virtual Audio Stream” and a “Virtual eText Stream.” (EX-1001 (’907 patent), 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.)

B. Claims

36. I have been asked to consider claims 1-21 in this Declaration. Of those, claims 1 and 21 are independent. Claim 1 is representative and recites a method comprising the following steps:

[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;

[b] storing location information for the plurality of digital media streams in the descriptor file;

[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;

[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;

[e] identifying synchronization points in the digital media content of the one or more other digital media streams;

[f] selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets; and

[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.

VI. CLAIMS 1-21 OF THE '907 PATENT WOULD HAVE BEEN OBVIOUS.

A. Claims 1-14, 17-19, and 21 Would Have Been Obvious in View of DTB.

37. For at least the reasons I discuss below, DTB renders claims 1-14, 17-19, and 21 obvious. DTB describes methods by which the text of a book can be “synchronized with the audio presentation” using a SMIL file. (EX-1003 (DTB), vii, 21.)

1. Claim 1

a. Preamble

38. The preamble recites “[a] method comprising.” DTB discloses the preamble at least because it discloses a method of creating digital talking books using SMIL files. (EX-1003 (DTB), Abstract, 6.)

b. Element 1[a][i]: Descriptor File

39. Element 1[a][i] recites “creating a descriptor file for synchronizing a plurality of digital media streams.” DTB discloses this claim element.

40. 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 (DTB), 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 (disclosing creating DTB files, which include SMIL files), 7-9, 12 (producer creates a DTB).)

41. DTB therefore discloses creating a descriptor file (SMIL file) for synchronizing a plurality of digital media streams (text, audio, and image streams). Thus, DTB discloses claim element 1[a][i].

**c. Elements 1[a][ii]-1[a][iii]: Digital Media Streams
Corresponding to a Same Originating Work**

42. Elements 1[a][ii] and 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.” DTB discloses these claim elements.

43. The '907 patent explains that “[a]n Originating Work is the complete text of a printed document, such as an article, speech, book or other written work[.]” (EX-1001 ('907 patent), 19:40-46, 19:56-60.)

44. 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 (DTB), 6-9, 12 (describing a “narrator” attribute that includes the “[n]ame of the person whose recorded voice is embodied in the DTB”), 66 (the SMIL file “permit[s] concurrent display of the text being narrated”), 60-61 (DTB may include “full

audio”), vii, Abstract (format useful for reference works, text books, and magazines), vii, 4.)

45. DTB also discloses that the plurality of media streams contains at least one other digital media stream, such as the complete “text” and/or “images” of the originating work (e.g., book). (EX-1003 (DTB), 6 (digital talking books include “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”).)

46. Accordingly, DTB discloses that the plurality of digital media streams (e.g., text, image, audio 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). Thus, DTB discloses claim elements 1[a][ii] and 1[a][iii].

d. Element 1[a][iv]: Descriptor File External to the Audio Stream

47. Element 1[a][iv] recites “wherein the descriptor file is external to the first digital media stream.” DTB discloses this claim element.

48. I note that this limitation requires the descriptor file to be external to the audio file(s). (See ¶¶42-46, above.)

49. As shown below, DTB’s SMIL file (e.g., descriptor file) (highlighted) 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_std.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 (DTB), 14-15; *id.*, 27 (audio streams are “pointed to” by the SMIL files).)

50. Thus, DTB discloses claim element 1[a][iv].

e. Element 1[b]: Location Information

51. Element 1[b] recites “storing location information for the plurality of digital media streams in the descriptor file.” DTB discloses or renders obvious this claim element.

52. As illustrated below, 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 (DTB), 25-27; *see also id.*, 98, 19, 3, 32.)

53. DTB’s URI is a locator because it “[p]oints to,” for example, a “segment of audio content to be rendered.” (EX-1003 (DTB), 25-27.) Indeed, DTB discloses that the “src” attribute “specifies by URI *the location of*” the file. (*Id.*, 98 (emphasis added).) This would have been consistent with a person of ordinary skill in the art’s understanding that src attributes in SMIL files are “used for locating and fetching the associated media.” (EX-1051 (SMIL Standard), 108; EX-1007 (Bulterman), 90-91 (“[t]he src (source) attribute gives the location of a media item”), 86; *see* EX-1049 (RFC 2396), 2.)

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

55. Alternatively, I note that if “location information” includes a location within the file (e.g., play location), DTB discloses this claim element 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 (DTB), 21-24, 3, 6-7, 27.)

56. Thus, DTB discloses or renders obvious claim element 1[b].

f. Element 1[c][i]: Time Offsets

57. Element 1[c][i] recites “identifying a plurality of time offsets in a time-line of the digital audio narration of the first digital media stream.” DTB discloses or renders obvious this claim element.

58. I note that the '907 patent explains that “time offset” “refers to the period of time, or difference in time, between two points in time[.]” (EX-1001 ('907 patent), 21:60-62.)

59. DTB’s audio “files are indexed by time-based positions” in the SMIL file. (EX-1003 (DTB), 21.) These time points 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.) The offsets are “identified” at least when they are indexed or otherwise determined by the DTB producer. (*Id.*, 21.)

60. These offsets are also “in a timeline of the digital audio narration of the first digital media stream,” as recited in claim element 1[c][i]. In instances where DTB’s audio stream is a single file, the offsets represent the time from the beginning of the audio stream to the selected point. (EX-1003 (DTB), 37 (an audio file ordinarily contains the entire content of the DTB).) When the audio stream comprises

multiple audio files, the time offset represents the time from the beginning of a particular audio file. (*Id.*, 7 (audio files for large books may span several media units).) Because each file in the audio stream may represent its own timeline of the audio stream, any time offsets within that audio file are time offsets in *a* timeline of the first digital media stream.

61. Accordingly, DTB discloses or renders obvious identifying (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). Thus, DTB discloses or renders obvious claim element 1[c][i].

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

62. Element 1[c][ii] recites “wherein the plurality of time offsets correspond to a plurality of content points in the digital audio narration.” DTB discloses this claim element.

63. I note that the ’907 patent specification refers to the beginning of a chapter, paragraph, or sentence as examples of “points” in the digital media stream. (EX-1001 (’907 patent), 21:42-45.)

64. Media content is grouped within the SMIL file into “parallels” or <par> elements. (EX-1003 (DTB), 26, 27 (describing a <par> element as a “[p]arallel time grouping in which multiple elements (e.g., text, audio, and image) play back

simultaneously”).) Each parallel comprises portions of media streams that comprise the same “body of material (e.g., the same paragraph)” or heading. (*Id.*, 29-30, 26, 32.) These bodies of material (e.g., paragraphs or headings) are content points. For example, the below image of a SMIL file shows a <par> for two **content points** (**green**): a heading (abbreviated “**h1**”) and a paragraph (abbreviated “**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>
```

(*Id.*, 32.)

65. Accordingly, DTB discloses that the plurality of time offsets (e.g., clip-Begin and clipEnd) correspond to a plurality of content points (e.g., headings and/or paragraphs) in the digital audio narration. Thus, DTB discloses claim element 1[c][ii].

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

66. 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 discloses this claim element.

67. DTB's content points (e.g., h1, para1, etc.) are identified in the parallel's id attribute. (EX-1003 (DTB), 32.) DTB's plurality of time offsets (e.g., clipBegin and clipEnd attributes) (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>
```

(*Id.*; see also *id.*, 26.)

68. Accordingly, DTB discloses storing the plurality of time offsets (e.g., clipBegin and/or clipEnd attributes) and the plurality of content points (e.g., h1, para1, etc. as reflected in parallel id attributes) in the descriptor file in a manner indicating a correlation between them. Thus, DTB discloses claim element 1[d].

i. Element 1[e]: Synchronization Points

69. Element 1[e] recites “identifying synchronization points in the digital media content of the one or more other digital media streams.” DTB discloses or renders obvious this claim element.

70. I note that the '907 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 ('907 patent), 31:4-17.)

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

72. For example, the SMIL file reproduced below shows several synchronization points, highlighted in teal. (EX-1003 (DTB), 32.) The parallel with id attribute “h1” 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 parallel with id attribute “para1” 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 parallel with id attribute “img1” 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.*)

Amazon.com, Inc. v. Audio Pod IP, LLC
Declaration of Ketan Mayer-Patel – U.S. Pat. No. 9,729,907

```
<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">
  
<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.*)

73. The synchronization points are determined by a DTB producer.
(EX-1003 (DTB), 21.)

74. Accordingly, DTB discloses or renders obvious identifying (e.g., determining) 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). Thus, DTB discloses or renders obvious claim element 1[e].

j. Element 1[f]: Synchronization Time Offsets

75. Element 1[f] recites “selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets.” DTB discloses or renders obvious this claim element.

76. DTB discloses that “[w]here text renditions of a segment are present, they are synchronized with the corresponding audio ... in the SMIL files[.]” (EX-1003 (DTB), 61.) When a text or image segment is available to be 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.” 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 id.*, 60-61.) Moreover, because “[t]he DTB producer determines granularity of the synchronization events,” the synchronization time offsets are “selected.” (*Id.*, 21)

77. Accordingly, DTB discloses or renders obvious selecting (e.g., determining) synchronization time offsets (e.g., clipBegin and/or clipEnd attributes contained in parallels with corresponding text segments or images) that correspond to the synchronization points (e.g., h1_1, para_1, caption_1) from the plurality of time offsets (e.g., all available clipBegin and clipEnd attributes). Thus, DTB discloses or renders obvious claim element 1[f].

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

78. 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.” DTB discloses this claim element.

79. 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 such that the parallel indicates a correlation between the synchronization time offsets and the synchronization points:

```
<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">
  
  <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 (DTB), 32.)

80. Thus, DTB discloses claim element 1[g][i].

I. Element 1[g][ii]: Synchronized Rendering

81. 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 discloses this claim element.

82. DTB’s SMIL (e.g., descriptor) file “provides for th[e] parallel and synchronized presentation” of audio, text, and images on DTB players (e.g., PC-based playback systems). (EX-1003 (DTB), 21; *see also id.*, 3, 6-7, 61 (“Text, audio, and any images are synchronized in the SMIL files.”); *see* ¶¶39-41, above (SMIL file is descriptor file for synchronizing plurality of digital media streams).)

83. Thus, DTB discloses claim element 1[g][ii].

84. Accordingly, for at least the reasons I discuss above, DTB renders claim 1 obvious as a whole.

2. Claim 2

85. 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 or renders obvious the additional claim element of claim 2.

86. DTB suggests delivery of content, including a descriptor file, from servers accessible to a client device because it discloses playback of multimedia presentations “over the Internet,” which a person of ordinary skill in the art would

have understood involves retrieving content that was not originally resident on the client device. (EX-1003 (DTB), 21.) It would have been obvious to a person of ordinary skill to store the descriptor file on a server accessible to the client device for multiple reasons.

87. First, storing a SMIL file on a client accessible server was widely known and commonly implemented in 2005. (*E.g.*, EX-1086 (Shteyn), 3:33-34 (“[c]ommunicating with a remote server is a well known technology”); EX-1032 (Leighton), 4:57-67 (“client machine is connected to a Web server via a network”).)

88. Second, a person of ordinary skill would have understood that storing the SMIL file on a client accessible server would have provided many benefits, including providing the ability to update SMIL files to, for example, indicate the best cache server location, which in turn improves streaming experience. (EX-1006 (Yoshimura), 1779 (portal server updates SMIL file to indicate best cache server locations), 1781 (updates “enable faster pre-fetching” and “lead[] to low latency”).) It would also allow for centralized management of SMIL files, controlling access to the SMIL file, and version control for updates.

89. A person of ordinary skill 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 person of ordinary skill. (*See, e.g.*, EX-1086 (Shteyn), 3:19-4:2; EX-1102 (Elsen), 50; EX-1107 (Griffin) ¶[0013].)

90. Thus, DTB discloses or renders obvious the additional claim element of claim 2.

91. Accordingly, for at least the reasons I discuss above, DTB renders claim 2 obvious as a whole.

3. Claim 3

92. 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 the additional claim element of claim 3 for the reasons I discuss above, including that DTB’s SMIL file enables synchronized rendering of a plurality of digital media streams on a client device. (*See, e.g.*, EX-1003 (DTB), 7; *see* ¶¶81-83, above.)

93. Accordingly, for at least the reasons I discuss above, DTB renders claim 3 obvious as a whole.

4. Claim 4

94. 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 discloses or renders obvious the additional claim elements of claim 4.

95. Digital talking book players “use the synchronization information [in the SMIL file] to ... track, during audio playback, the corresponding position in the textual content file[.]” (EX-1003 (DTB), 7.) This “tracked” information from the SMIL file 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.)

96. 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) that contains the <bookmark> or <lastmark> position (red) and (2) a time offset to the exact point identified in the <bookmark> or <lastmark> element (e.g., bookmarked position) (blue). (EX-1003 (DTB), 47-48.)

Amazon.com, Inc. v. Audio Pod IP, LLC
Declaration of Ketan Mayer-Patel – U.S. Pat. No. 9,729,907

```
<title>
  <text>Gone with the Wind</text>
  <audio src="gtw_title.mp3" />
</title>

<uid>us-rfbd-JT065</uid>

<lastmark>
  <ncxRef>gtw.ncx#lvl1_5</ncxRef>
  <uri>gtw_ch5.smil#para023</uri>
  <timeOffset>173</timeOffset>
</lastmark>

<bookmark>
  <ncxRef>gtw.ncx#lvl1_1</ncxRef>
  <uri>gtw_ch1.smil#para008</uri>
  <timeOffset>22</timeOffset>
</bookmark>
```

(*Id.*, 49.) The position may also be identified by a character offset. (*Id.*, 47, 50.)

97. Accordingly, DTB discloses or renders obvious tracking a current position (e.g., by the playback device using the SMIL file) 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 (e.g., audio, text, and image streams) 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 offset) and identifies the descriptor file (e.g., URI pointing to SMIL file). Thus, DTB discloses or renders obvious the additional claim element of claim 4.

98. Accordingly, for at least the reasons I discuss above, DTB renders claim 4 obvious as a whole.

5. Claim 5

99. 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 discloses or renders obvious the additional claim element of claim 5.

100. Digital talking book players can resume play at either the <bookmark> or <lastmark> position on the playback device. (See ¶¶94-98, above; EX-1003 (DTB), 47-48; *id.*, 48 (bookmark is “for direct access in future”), 45-50.) Moreover, because bookmark files are “exportable,” it would have been obvious to a person of ordinary skill in the art to use them to resume playback from the bookmarked position(s) on a second other client device. (EX-1003 (DTB), 7; *see* EX-1107 (Griffin) ¶[0005].) Thus, DTB discloses or renders obvious the additional claim element of claim 5.

101. Accordingly, for at least the reasons I discuss above, DTB renders claim 5 obvious as a whole.

6. Claim 6

102. 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 discloses or renders obvious the additional claim element of claim 6.

103. DTB bookmarks are “based on the audio file *or* the textual content file.” (EX-1003 (DTB), 45 (emphasis added).) 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. Further, DTB provides examples of bookmarks, including a <bookmark> element set using an audio stream (but not using the text stream) and a <lastmark> element set using a text stream (but not the audio stream):

```
<bookmark>
  <ncxRef>gwtw.ncx#lv11_1</ncxRef>
  <uri>gwtw_ch1.smil#para008</uri>
  <timeOffset>22</timeOffset>
</bookmark>

<lastmark>
  <ncxRef>chemtd.ncx#lv11_3</ncxRef>
  <uri>chemtd.xml#para297</uri>
  <charOffset>130</charOffset>
</lastmark>
```

(*Id.*, 49; *id.*, 50.) Thus, DTB discloses or renders obvious the additional claim element of claim 6.

104. Accordingly, for at least the reasons I discuss above, DTB renders claim 6 obvious as a whole.

7. Claim 7

a. Element 7[a]: Visually Rendered Content

105. Claim 7 depends from claim 3. 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 claim element for the reasons I discuss above. (EX-1003 (DTB), 6, vii; *see* ¶¶42-46, above.)

b. Element 7[b]: Automated Page Turns

106. 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 or renders obvious this claim element.

107. DTB discloses this limitation because it discloses that “page numbers” can be “presented ... during normal playback of a DTB.” (EX-1003 (DTB), 25; *see also id.*, 95 (page number from the print document is “inserted at the point within the file immediately preceding the first item of content on a new page” and describing page numbering schemes).)

108. Accordingly, 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 sequential playback of a DTB). Thus, DTB discloses or renders obvious claim element 7[b].

109. Accordingly, for at least the reasons I discuss above, DTB renders claim 7 obvious as a whole.

8. Claim 8

110. 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 or renders obvious the additional claim element of claim 8.

111. As I discuss above, DTB discloses selecting synchronization time offsets down to the “paragraph, sentence, or even word level.” (*See* ¶¶69-80, above.) A person of ordinary skill in the art 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). Thus, DTB discloses or renders obvious the additional claim element of claim 8.

112. Accordingly, for at least the reasons I discuss above, DTB renders claim 8 obvious as a whole.

9. Claim 9

113. 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 discloses or renders obvious the additional claim element of claim 9.

114. As illustrated below, DTB’s audio file (for example, “rs_fwdx.mp3”) may be divided into audio “segment[s]” or “clip[s]” that are represented as <audio> elements in the SMIL file 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 (DTB), 31-34, 27.) DTB also provides examples where the audio segments defined by the offsets comprise separate audio files. (*Id.*, 33-34, 6 (narration embodied in multiple audio files), 15 (manifest disclosing multiple audio files in a single DTB).)

115. Accordingly, 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). Thus, DTB discloses or renders obvious the additional claim element of claim 9.

116. Accordingly, for at least the reasons I discuss above, DTB renders claim 9 obvious as a whole.

10. Claim 10

117. 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.” DTB discloses the additional claim element of claim 10.

118. The clipBegin and clipEnd times reflect the start and end time of the content segment identified in each parallel. (EX-1003 (DTB), 27, 31-34.) Thus, DTB discloses the additional claim element of claim 10.

119. Accordingly, for at least the reasons I discuss above, DTB renders claim 10 obvious as a whole.

11. Claim 11

a. Element 11[a]: Selecting Two or More of the Streams

120. Claim 11 depends from claim 1. Element 11[a] recites “selecting two or more of the plurality of digital media streams.” DTB discloses or renders obvious this claim element.

121. A digital talking book user selects media streams by selecting a book and/or jumping to a particular section of the book (e.g., a chapter, heading, section, page, figure, etc.) using bookmarks or lastmarks. (EX-1003 (DTB), 55, 41.) In

response, the player selects two or more of the digital media streams (e.g., audio and text) for playback. (*Id.*, 61 (describing books having audio and text), 62 (content rendering).) Thus, DTB discloses or renders obvious claim element 11[a].

b. Element 11[b]: Determining a First Position

122. 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 or renders obvious this claim element because it discloses determining a position using the synchronization time offsets both (1) during playback and (2) to access a bookmark/lastmark position.

123. First, DTB discloses that during playback, digital talking book players use the synchronization information including the offsets in the audio file to track the corresponding position in the text and image content files. (EX-1003 (DTB), 7 (“DTB players use the synchronization information ... to track, during audio playback, the corresponding position in the textual content file”); *see* ¶¶94-98, above.)

124. Second, to access a marked location, 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 of the SMIL file. (EX-1003 (DTB), 46-50 (“the reader has set two bookmarks, one in chapter 1, 22 seconds from the start of paragraph 8, and the other in chapter 3, 88 seconds from the start of paragraph 12”).)

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). 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. (*Id.*, 6-7 (“DTB players use the synchronization information to both access points in the audio presentation and to track, during audio playback, the corresponding position in the textual content file”), 21-27, 60-61.)

125. Thus, DTB discloses or renders obvious claim element 11[b].

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

126. 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 this claim element.

127. For at least the reasons I discuss above, 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). (*See*

¶¶81-83 (SMIL file allows synchronized rendering of audio and text on client device), 122-125, 92-93 (SMIL file is used to synchronize plurality of digital media streams).) Thus, DTB discloses claim element 11[c].

128. Accordingly, for at least the reasons I discuss above, DTB renders claim 11 obvious as a whole.

12. Claim 12

129. 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 discloses or renders obvious the additional claim element of claim 12.

130. DTB users can “step through and choose from a group of” bookmarks to initiate playback from (and therefore determine) a particular location. (EX-1003 (DTB), 45-50.) Accordingly, 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). Thus, DTB discloses or renders obvious the additional claim element of claim 12.

131. Accordingly, for at least the reasons I discuss above, DTB renders claim 12 obvious as a whole.

13. Claim 13

132. 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 or renders obvious the additional claim element of claim 13 for the reasons I discuss above. (*See* ¶¶94-98, 120-128, above.)

133. Accordingly, for at least the reasons I discuss above, DTB renders claim 13 obvious as a whole.

14. Claim 14

134. Claim 14 depends from claim 1 and further recites “wherein the descriptor file contains information that is human readable when rendered.” DTB discloses or renders obvious the additional claim element of claim 14.

135. 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, as I discuss above. (*See* ¶¶51-56, above.)

136. 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.) Accordingly, 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. Thus, DTB discloses or renders obvious the additional claim element of claim 14.

137. Accordingly, for at least the reasons I discuss above, DTB renders claim 14 obvious as a whole.

15. Claim 17

138. 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 the additional claim element of claim 17 for the reasons I discuss above. (*See* ¶¶42-46, above.)

139. Accordingly, for at least the reasons I discuss above, DTB renders claim 17 obvious as a whole.

16. Claim 18

a. Element 18[a][i]: Navigating

140. Claim 18 depends from claim 1. 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 or renders obvious this claim element.

141. DTB discloses “local navigation” that “can be controlled by the ... SMIL file(s).” (EX-1003 (DTB), 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* ¶¶94-98, 99-101 (disclosing navigating using bookmarks).) 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). Even if DTB did not disclose this limitation, it would have been obvious to a person of ordinary skill in the art to store the SMIL file on the client device for “easy access by the player.” (*See* EX-1003 (DTB), 15, 55.) Thus, DTB discloses or renders obvious claim element 18[a][i].

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

142. 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 or renders obvious this claim element.

143. DTB suggests delivery of content from servers to a client device because it discloses playback of multimedia presentations “over the Internet,” which a person of ordinary skill would have understood would involve retrieving content that was not originally resident on the client device. (EX-1003 (DTB), 21.)

144. Even if DTB did not disclose this limitation, it would have been obvious to a person of ordinary skill. Such a person 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. In addition, a person of ordinary skill would have understood content was either stored locally or remotely, both of which would have been obvious.

145. Thus, DTB discloses or renders obvious claim element 18[a][ii].

146. Accordingly, for at least the reasons I discuss above, DTB renders claim 18 obvious as a whole.

17. Claim 19

147. 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 or renders obvious the additional claim element of claim 19.

148. The ’907 patent explains that “an example of a discontinuous media stream is an e-Book downloaded as a plurality of files.” (EX-1001 (’907 patent), 19:63-64.)

149. DTB discloses multiple examples of digital talking books with discontinuous media streams (e.g., media streams comprising multiple files). For example, DTB’s text stream can be discontinuous. (EX-1003 (DTB), 17 (“a DTB can contain multiple textual content files”).) Similarly, DTB discloses discontinuous audio streams. (*Id.*, 2, 15 (manifest illustrating multiple audio files); *see also id.*, 61.) Accordingly, DTB discloses or renders obvious a discontinuous digital media stream (e.g., text and/or audio stream) that is discontinuous when rendered. Thus, DTB discloses or renders obvious the additional claim element of claim 19.

150. Accordingly, for at least the reasons I discuss above, DTB renders claim 19 obvious as a whole.

18. Claim 21

151. Independent claim 21 recites substantially the same limitations as claim 1, but 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 method steps of claim 1.

152. DTB explains that producers of digital talking books can use computers with PC-based browsers and uses “software” to “generate the SMIL file.” (See EX-1003 (DTB), 30-31, 3, 60 (DTB uses PC-based browsers and playback systems), 15 (text streams are computer readable formats), 55-56 (discussing using a hard drive).) Thus, DTB discloses or renders obvious claim 21’s preamble.

153. I have compared the elements in claim 21 to the elements in claim 1. As shown, the actions recited in claim 21 mirror the method steps recited in claim 1. The table below demonstrates how the elements correspond. As I discuss above for claim 1, DTB discloses or renders obvious the actions recited in claim 21. (See ¶¶38-84, above.)

	Claim 21		Claim 1
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:	1[pre]	A method comprising:
21[a]	<i>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</i>	1[a]	<i>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</i>

	Claim 21		Claim 1
	<i>the originating work and one or more other digital media streams, and wherein the descriptor file is external to the first digital media stream;</i>		<i>the originating work and one or more other digital media streams, and wherein the descriptor file is external to the first digital media stream;</i>
21[b]	<i>store location information for the plurality of digital media streams in the descriptor file;</i>	1[b]	<i>storing location information for the plurality of digital media streams in the descriptor file;</i>
21[c]	<i>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;</i>	1[c]	<i>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;</i>
21[d]	<i>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;</i>	1[d]	<i>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;</i>
21[e]	<i>identify synchronization points in the digital media content of the one or more other digital media streams;</i>	1[e]	<i>identifying synchronization points in the digital media content of the one or more other digital media streams;</i>
21[f]	<i>select synchronization time offsets that correspond to the synchronization points from the plurality of time offsets; and</i>	1[f]	<i>selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets; and</i>
21[g]	<i>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</i>	1[g]	<i>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</i>

	Claim 21		Claim 1
	<i>synchronization points, such that the descriptor file allows a synchronized rendering of the plurality of digital media streams on a client device.</i>		<i>the synchronization points, such that the descriptor file allows a synchronized rendering of the plurality of digital media streams on a client device.</i>

154. Accordingly, for at least the reasons I discuss above, DTB renders claim 21 obvious as a whole.

B. Claims 2 and 18-19 Would Have Been Obvious in View of DTB and Yoshimura.

155. As I discuss above, DTB discloses or renders obvious each limitation of claims 2 and 18-19. (See ¶¶85-91, 140-150, above.) For at least those reasons and the reasons I discuss below, claims 2, 18, and 19 would also have been obvious in further view of Yoshimura. I incorporate my above discussion of DTB here, and only discuss claim elements for which Yoshimura is relevant in combination with DTB.

1. Claim 2

156. Claim 2 depends from claim 1 and further recites “storing the descriptor file on a server that is accessible to the client device.” Yoshimura discloses the additional claim element of claim 2.

157. 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 (Yoshimura), 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 (describing modification of SMIL files at the portal server), 1782 (“the portal server returns the modified SMIL file to the mobile client”).) Thus, Yoshimura discloses the additional claim element of claim 2.

158. A person of ordinary skill in the art would have been motivated to store DTB’s SMIL file on a server, as taught by Yoshimura, for several reasons.

159. First, a person of ordinary skill would have been motivated to do so for the reasons I discuss above. (*See* ¶¶85-91, above.)

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

161. Third, 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 accessible to the client device). Fourth, 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. Fifth, 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).

162. A person of ordinary skill would have reasonably expected success for the reasons I discuss above, and because both references disclose using the same file type (SMIL file) for the same purpose (synchronization of audio and text data). (*See* ¶¶85-91, above; EX-1003 (DTB), 21; EX-1006 (Yoshimura), 1779-80.)

163. Accordingly, for at least the reasons I discuss above, DTB and Yoshimura render claim 2 obvious as a whole.

2. Claim 18

a. Element 18[a][i]: Navigating

164. Claim 18 depends from claim 1. 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.” Yoshimura discloses this claim element because Yoshimura discloses that the client downloads SMIL files from a server and therefore stores them on the client once downloaded. (EX-1006 (Yoshimura), 1779, 1782, 1785.)

165. A person of ordinary skill in the art would have been motivated to store the descriptor file (e.g., SMIL file) on the device as disclosed in Yoshimura for several reasons.

166. First, a person of ordinary skill would have been motivated to do so for the reasons discussed above. (*See* ¶¶140-146, above.)

167. Second, 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). Third, 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. Fourth, 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).

168. Finally, methods involving transferring descriptor files, including SMIL files, to a client were commonplace and disclosed by numerous references, giving a person of ordinary skill a reasonable expectation of success in the modification. (*See, e.g.*, EX-1086 (Shteyn), 3:19-31 ("the client contacts the server, selects the particular content file and downloads the control information" which is "coded in Extensible Markup Language (XML)"); EX-1029 (Steele), 3; EX-1106 (Chen), 888-89.)

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

169. 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.” Yoshimura discloses this claim element.

170. Yoshimura’s clients use information in SMIL files to “receive multi-media 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 (Yoshimura), 1778-79, 1782.) Accordingly, 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). Thus, Yoshimura discloses this claim element.

171. A person of ordinary skill would have been motivated to combine the teachings of DTB and Yoshimura, and to retrieve content from a server for several reasons.

172. First, a person of ordinary skill would have been motivated to do so for the reasons I discuss above. (See ¶¶140-146, 156-163, 164-168, above.)

173. Second, 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). Third, 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. Fourth, 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).

174. Fifth, 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 person of ordinary skill.

175. Finally, a person of ordinary skill would have reasonably expected success in this combination because retrieving content from a remote source was commonplace and disclosed by numerous references. (*See, e.g.*, EX-1086 (Shteyn), 3:19-4:2; EX-1032 (Leighton), 17:4-13; EX-1009 (Copley); EX-1029 (Steele), 4).)

176. Thus, for at least the reasons I discuss above, DTB and Yoshimura render claim 18 obvious as a whole.

3. Claim 19

177. 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.” Yoshimura discloses the additional claim element of claim 19 because Yoshimura discloses discontinuous audio and video streams wherein the individual files are fetched (and rendered) separately. (EX-1006 (Yoshimura), 1780-82; *see* ¶¶156-176, above.) It would have been obvious to combine DTB and Yoshimura to achieve the claimed results for the reasons I discuss above. (*See* ¶¶156-176, 147-150, above.)

178. Accordingly, for at least the reasons I discuss above, DTB and Yoshimura render claim 19 obvious as a whole.

C. Claim 7 Would Have Been Obvious in View of DTB and Duncan.

179. As I discuss above, DTB discloses or renders obvious each limitation of claim 7. (*See* ¶¶105-109, above.) For at least those reasons and the reasons I discuss below, claim 7 would also have been obvious in further view of Duncan. I incorporate my above discussion of DTB here, and only discuss claim elements for which Duncan is relevant in combination with DTB. In particular, DTB discloses claim element 7[a] for the reasons I discuss above. (*See* ¶105, above.)

180. Element 7[b] recites “providing automated page turns as the two or more of the plurality of digital media streams are simultaneously rendered.” Duncan discloses this claim element.

181. Duncan discloses a digital talking book (EX-1027 (Duncan), Abstract) where both graphics and audio are presented (*id.* ¶[0033]). Duncan discloses that,

when the 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.*) Thus, Duncan discloses this claim element.

182. It would have been obvious to a person of ordinary skill in the art 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).

183. A person of ordinary skill would have been motivated to combine Duncan’s automatic page turning with DTB’s player for several reasons.

184. First, both references seek to “accommodate sight-impaired readers.” (EX-1027 (Duncan) ¶[0004]; EX-1003 (DTB), vii.) It would have been obvious to a person of ordinary skill 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 (DTB), Abstract; EX-1027 (Duncan) ¶¶[0004], [0042].)

185. 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). 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. For example, Duncan's automatic page turn feature is accomplished by processing character offsets in an XML text stream (EX-1027 (Duncan) ¶¶[0041]-[0042]), which DTB also discloses. (EX-1003 (DTB), 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).

186. Finally, methods to automate page turning were commonplace and disclosed by numerous references, giving a person of ordinary skill a reasonable expectation of success in the modification. (*See, e.g.*, EX-1059 (Hendricks), 49:6-7; EX-1057 (Takahashi), Abstract; EX-1058 (Fujisawa), 4:32-40.)

187. Accordingly, for at least the reasons I discuss above, DTB and Duncan render claim 7 obvious as a whole.

D. Claim 8 Would Have Been Obvious in View of DTB and Heckerman.

188. As I discuss above, DTB discloses or renders obvious each limitation of claim 8. (*See* ¶¶110-112, above.) For at least those reasons and the reasons I

discuss below, claim 8 would also have been obvious in further view of Heckerman.

I incorporate my above discussion of DTB here.

189. 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.” Heckerman discloses the additional claim element of claim 8.

190. 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 (Heckerman), 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.*) Accordingly, 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. Thus, Heckerman discloses the additional claim element of claim 8.

191. A person of ordinary skill in the art 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.

192. First, a person of ordinary skill 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.

193. Second, a person of ordinary skill would have been “motivated to do so in order to exploit acoustic cues to segment recorded speech into semantically meaningful chunks,” as the Patent Office determined during examination of a related patent application. (EX-1074 (’933 Final Office Action), 5.)

194. 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). 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. 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 (Heckerman), 3:6-10; EX-1003 (DTB), 21 (describing use of SMIL files to synchronize audio and text to create a digital talking book).) 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).

195. 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 (EP '154 File History), 5), and disclosed by numerous references. Thus, a person of ordinary skill would have had a reasonable expectation of success in the modification. (*See, e.g.*, EX-1069 (Dimitrova), 2:4-6; EX-1063 (Delacourt), 113-15; EX-1075 (Arons), 9-10.)

196. Accordingly, for at least the reasons I discuss above, DTB and Heckerman render claim 8 obvious as a whole.

E. Claims 14 and 20 Would Have Been Obvious in View of DTB and Bulterman.

197. As I discuss above, DTB discloses or renders obvious each limitation of claim 14. (*See* ¶¶134-137, above.) For at least those reasons and the reasons I discuss below, claim 14 would also have been obvious in further view of Bulterman. I incorporate my above discussion of DTB here. Furthermore, claim 20 depends from claim 1. As I discuss above, DTB discloses or renders obvious each limitation of claim 1. (*See* ¶¶38-84, above.) For at least those reasons and the reasons I discuss below, claim 20 would have been obvious in further view of Bulterman.

1. Claim 14

198. Claim 14 further recites “wherein the descriptor file contains information that is human readable when rendered.” Bulterman discloses the additional claim element of claim 14.

199. Bulterman teaches embedding text directly in a SMIL file. (EX-1007 (Bulterman), 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 the additional claim element of claim 14.

200. This was also disclosed in many other references. (*See, e.g.*, EX-1101 (Simpson), 2 (“literal text” can be included in the SMIL file using a special form of

the “src” attribute); EX-1103 (Shane), 6-7 (SMIL uses “in-line text instead of referring to separate [sic] plain-text files as the text source.”).)

201. A person of ordinary skill in the art would have been motivated to directly embed text in DTB’s SMIL file for several reasons.

202. First, a person of ordinary skill 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. Indeed, Bulterman advises that “[f]or short text strings, this is a convenient mechanism[.]” (EX-1007 (Bulterman), 87.)

203. Second, a person of ordinary skill 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. Both would have been obvious to a person of ordinary skill and such a person would immediately have understood Bulterman’s inclusion of text in the SMIL file to be a viable method.

204. 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). 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. 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).

205. Finally, including text to be rendered within the SMIL file was commonplace and disclosed by numerous references, giving a person of ordinary skill a reasonable expectation of success in the modification. (*See, e.g.*, EX-1101 (Simpson), 2 (“literal text” can be included in the SMIL file using a special form of the “src” attribute); EX-1103 (Shane), 6-7 (SMIL file uses “in-line text instead of referring to separate [sic] plain-text files as the text source”); EX-1051 (SMIL Standard), 114 (explaining that for “user agents that cannot display a particular media object,” the alt attribute “specifies alternate text[,]” and that “[i]t is strongly recommended that all media object elements have an ‘alt’ attribute with a brief, meaningful description”); EX-1104 (Rutledge), 83 (explaining that “[d]escriptive constructs [like alt] enhance adaptivity by alerting disabled users to presentation content that they can’t directly perceive.... [u]sers who cannot perceive a document’s presented forms can then read... their descriptions instead.”).)

206. Accordingly, for at least the reasons I discuss above, DTB and Bulterman render claim 14 obvious as a whole.

2. Claim 20

207. 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.” Bulterman discloses the additional claim element of claim 20.

208. The “src” attribute specifies the storage location of digital media streams. (See ¶¶51-56, above; EX-1003 (DTB), 98 (the “src” attribute “specifies by URI the location of the image file”).) 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 (Bulterman), 86-87.) Thus, Bulterman discloses the additional claim element of claim 20.

209. Accordingly, for at least the reasons I discuss above, DTB and Bulterman render claim 20 obvious as a whole.

F. Claim 15 Would Have Been Obvious in View of DTB and Yang.

210. 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.” As I discuss above, DTB discloses or renders obvious each

limitation of claim 1. (*See* ¶¶38-84, above.) Yang discloses the additional claim element of claim 15.

211. Yang discloses a “just-in-time” retrieval system for synchronized multimedia presentations using SMIL files. (EX-1045 (Yang), 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.*) Thus, Yang discloses the additional claim element of claim 15.

212. A person of ordinary skill in the art would have been motivated to combine the teachings of DTB and Yang, and to deliver DTB’s content with Yang’s just-in-time method for several reasons.

213. 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 (Yang), 49.)

214. 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). 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. The combination would improve the user experience because it would result in a smoother streaming session. (EX-1006 (Yoshimura), 1782.) 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).

215. Finally, methods involving prefetching segments were commonplace and disclosed by numerous references, giving a person of ordinary skill a reasonable expectation of success in the modification. (EX-1006 (Yoshimura), 1782-83; EX-1086 (Shteyn), 4:8-25 ("the next file segment is downloaded at the client and stored in a buffer while the previous file segment... is being played out"); EX-1105 (Shin), 577.)

216. Accordingly, for at least the reasons I discuss above, DTB and Yang render claim 15 obvious as a whole.

G. Claim 16 Would Have Been Obvious in View of DTB, Yang, and Copley.

217. 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." As I discuss above, DTB and Yang disclose or render obvious each

limitation of claim 15. (*See* ¶¶210-216, above.) Copley discloses the additional claim element of claim 16.

218. Copley discloses a streaming media system that uses “client-side performance monitoring to automatically detect and correct connection failures or performance problems.” (EX-1009 (Copley) ¶[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].) Thus, Copley discloses the additional claim element of claim 16.

219. A person of ordinary skill in the art 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.

220. 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 (Copley) ¶¶[0013]-[0015].)

221. 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). 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. The combination would improve the user experience because the streaming content would suffer fewer interruptions. 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).

222. Finally, server optimization based on failure conditions, network degradation, and throughput was commonplace and disclosed by numerous references, giving a person of ordinary skill a reasonable expectation of success in the modification. (*See, e.g.*, EX-1032 (Leighton), 12:53-67; EX-1029 (Steele), 4-5; EX-1006 (Yoshimura), 1781-82.)

223. Accordingly, for at least the reasons I discuss above, DTB, Yang, and Copley render claim 16 obvious as a whole.

H. Claims 1-3, 8-12, 14, 17-19, and 21 Would Have Been Obvious in View of McCartney.

224. For at least the reasons I discuss below, McCartney renders claims 1-3, 8-12, 14, 17-19, and 21 obvious.

1. Claim 1

a. Preamble

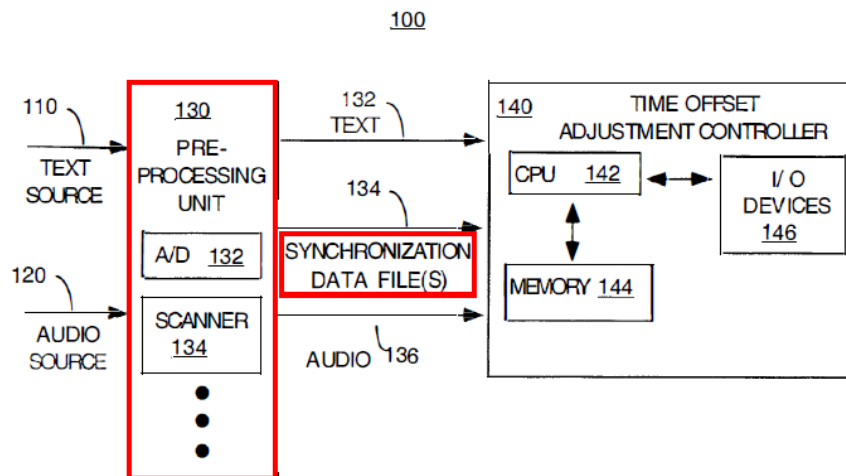
225. The preamble recites “[a] method comprising.” McCartney discloses or at least renders the preamble obvious because it discloses a “[m]ethod for constructing a digital talking book from text data and audio data.” (EX-1013 (McCartney), claim 1.)

b. Element 1[a][i]: Descriptor File

226. Element 1[a][i] recites “creating a descriptor file for synchronizing a plurality of digital media streams.” McCartney discloses or renders obvious this claim element.

227. 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 (McCartney), 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.)

228. 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).

229. Even if this limitation required a single descriptor file, it would have been obvious to a person of ordinary skill in the art in view of McCartney. It would have been obvious to a person of ordinary skill to combine McCartney’s BPM and TSD files’ data into a single file having characteristics of both the TSD and BPM files. A person of ordinary skill 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. For example, a person of ordinary skill would have understood that combining the files would simplify sharing and distribution because sending a

single merged file is more convenient than sending multiple separate files. Also, keeping related information in one file improves organization and makes access and referencing easier. A person of ordinary skill would have recognized that combining the files would save storage space and result in less redundancy and less chance for error or inconsistencies between the two documents. Combining the synchronization files would also result in more efficient data analysis at least because the system would only need to access a single file and would not need to process links between multiple files.

230. Moreover, a person of ordinary skill 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 be processed by “an XML application,” meaning they share a compatible format. (EX-1013 (McCartney) ¶¶[0026], [0031].)

231. Thus, McCartney discloses or renders obvious claim element 1[a][i].

c. Element 1[a][ii]: Same Originating Work

232. Element 1[a][ii] recites that “the plurality of digital media streams each contain digital media content corresponding to a same originating work.” McCartney discloses this claim element.

233. McCartney’s text files “contain the full text of the original printed book.” (EX-1013 (McCartney) ¶[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).

Thus, McCartney discloses claim element 1[a][ii].

d. Element 1[a][iii]: Audio Stream and Other Digital Media Stream

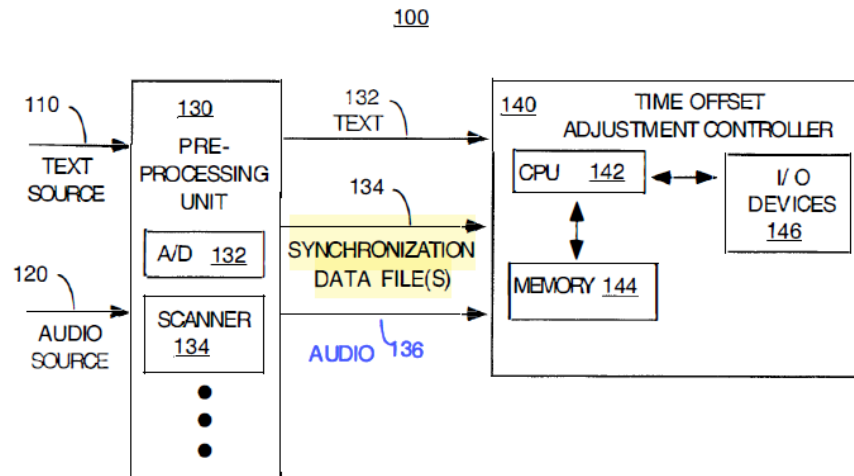
234. Element 1[a][iii] recites that “the plurality of digital media streams includes a first digital media stream containing a digital audio narration of the originating work and one or more other digital media streams.” McCartney discloses this claim element.

235. As I discuss above, 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 (McCartney) ¶¶[0020]-[0021]; *see* ¶¶226-233, above.) Thus, McCartney discloses claim element 1[a][iii].

e. Element 1[a][iv]: Descriptor File External to the Audio Stream

236. Element 1[a][iv] recites “wherein the descriptor file is external to the first digital media stream.” McCartney discloses this claim element.

237. 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 (McCartney), Fig. 1, ¶¶[0031]; *id.* ¶¶[0038]-[0039].) Thus, McCartney discloses claim element 1[a][iv].

f. Element 1[b]: Location Information

238. Element 1[b] recites “storing location information for the plurality of digital media streams in the descriptor file.” McCartney discloses or renders obvious this claim element because it discloses or renders obvious storing location information in its TSD and BPM files.

i. TSD File

239. 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. For example, a person of ordinary skill in the art would have

understood that files were located using a file path, part of which is the file name. Alternatively, it would have been obvious to a person of ordinary skill 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 (McCartney) ¶[0032].) A person of ordinary skill would have been motivated to include location information to assist in locating the referenced audio file.

240. 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 (McCartney) ¶¶[0032]-[0033], [0042]-[0043], Fig. 5.)

ii. BPM File

241. 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." (EX-1013 (McCartney) ¶[0029], 7.) 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.

242. 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 (McCartney) ¶[0030].)

243. Thus, McCartney discloses or renders obvious claim element 1[b].

g. Element 1[c][i]: Time Offsets

244. 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.” McCartney discloses this claim element.

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

ID	Start Time	End Time	Type
X pedal_0086	00:00:02.000	00:00:02.100	o
chap_0000	00:00:02.100	00:00:02.125	o
X pedal_0088	00:00:02.125	00:00:02.200	o
X pedal_0089	00:00:02.200	00:00:02.300	o
X pedal_0090	00:00:02.300	00:00:02.400	o
X pedal_0091	00:00:02.400	00:00:02.500	o
X pedal_0092	00:00:02.500	00:00:02.600	o
X pedal_0093	00:00:02.600	00:00:02.700	o
track6.wav			
X pedal_0094	00:00:00.000	00:00:00.100	o
X pedal_0095	00:00:00.100	00:00:00.200	o
X pedal_0096	00:00:00.200	00:00:00.300	o
X pedal_0097	00:00:00.300	00:00:00.400	o
chap_0001	00:00:00.400	00:00:00.450	o

(*Id.*, Fig. 5, ¶[0033].) McCartney’s TSD files, including the time offsets, are generated at the pre-processing unit. (*Id.* ¶[0023].)

246. Accordingly, McCartney discloses identifying a plurality of time offsets (e.g., start time and/or end time) in a timeline of the digital audio narration (audio file) of the first digital media stream (audio stream). Thus, McCartney discloses claim element 1[c][i].

h. Elements 1[c][ii] and 1[d]: Content Points in Audio Narration Correlated to Time Offsets

247. Element 1[c][ii] recites “wherein the plurality of time offsets correspond to a plurality of content points in the digital audio narration.” 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.” McCartney discloses or renders obvious these claim elements.

248. 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 (McCartney) ¶¶[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:

TIME STAMP DATA
(TSD) FILE 300

310 - DATA ELEMENT

- - 320 RECORD ELEMENT
- - 322 - ID
- - 324 - CLIP START
- - 326 - CLIP END
- - 328 - TYPE

ID	Start Time	End Time	Type
X pedal_0085	00:00:02.000	00:00:02.100	o
X chap_0000	00:00:02.100	00:00:02.125	o
X pedal_0088	00:00:02.125	00:00:02.200	o
X pedal_0089	00:00:02.200	00:00:02.300	o
X pedal_0090	00:00:02.300	00:00:02.400	o
X pedal_0091	00:00:02.400	00:00:02.500	o
X pedal_0092	00:00:02.500	00:00:02.600	o
X pedal_0093	00:00:02.600	00:00:02.700	o
track6.wav			
X pedal_0094	00:00:00.000	00:00:00.100	o
X pedal_0095	00:00:00.100	00:00:00.200	o
X pedal_0096	00:00:00.200	00:00:00.300	o
X pedal_0097	00:00:00.300	00:00:00.400	o
X chap_0001	00:00:00.400	00:00:00.450	o
X pedal_0098	00:00:00.450	00:00:00.500	o
X pedal_0100	00:00:00.500	00:00:00.600	o
X pedal_0101	00:00:00.600	00:00:00.700	o
X pedal_0102	00:00:00.700	00:00:00.800	o
X pedal_0103	00:00:00.800	00:00:00.900	o
X chap_0002	00:00:00.900	00:00:00.950	o

(Id., Figs. 5, 3.)

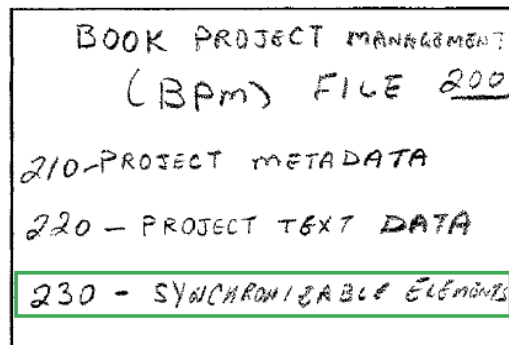
249. Accordingly, McCartney discloses or renders obvious storing the plurality of time offsets (e.g., clip start/end time attributes) and the plurality of content points (e.g., ID attributes) in the descriptor file (e.g., TSD file) in a manner indicating a correlation (e.g., in the same line as the same <record> elements) between the plurality of time offsets and the plurality of content points. (EX-1013 (McCartney) ¶¶[0033], Figs. 3, 5.) Thus, McCartney discloses claim elements 1[c][ii] and 1[d].

i. Element 1[e]: Synchronization Points

250. Element 1[e] recites “identifying synchronization points in the digital media content of the one or more other digital media streams.” McCartney discloses this claim element.

251. In McCartney, the “other digital media stream” is text presented, for example, in an HTML file. (EX-1013 (McCartney) ¶¶[0019]-[0020]; *see* ¶¶234-235, above.) McCartney links the text and audio files by chapters (sometimes referred to as “headings”) and/or pages. (EX-1013 (McCartney) ¶¶[0041]-[0043], [0046].) Each chapter 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.*; *see* ¶¶247-249 (ID attributes correspond to positions in the audio), above.)

252. Alternatively, McCartney’s BPM file lists “synchronizable elements within the source text files that are used to indicate points of navigation.” (EX-1013 (McCartney), 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.)

253. Accordingly, 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 (McCartney) ¶¶[0038], [0023].) Thus, McCartney discloses claim element 1[e].

j. Element 1[f]: Synchronization Time Offsets

254. Element 1[f] recites “selecting synchronization time offsets that correspond to the synchronization points from the plurality of time offsets.” McCartney discloses or renders obvious this claim element.

255. 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 (McCartney) ¶¶[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].) This is shown in Figure 5, which includes a plurality of content points (blue), a plurality of time

offsets (pink), linked synchronization points (green), and their corresponding synchronization time offsets (yellow):

ID	Start Time	End Time	Type
pedal_0086	00:00:02.000	00:00:02.100	o
chap_0000	00:00:02.100	00:00:02.125	o
pedal_0088	00:00:02.125	00:00:02.200	o
pedal_0089	00:00:02.200	00:00:02.300	o
pedal_0090	00:00:02.300	00:00:02.400	o
pedal_0091	00:00:02.400	00:00:02.500	o
pedal_0092	00:00:02.500	00:00:02.600	o
pedal_0093	00:00:02.600	00:00:02.700	o
track6.wav			
pedal_0094	00:00:00.000	00:00:00.100	o
pedal_0095	00:00:00.100	00:00:00.200	o
pedal_0096	00:00:00.200	00:00:00.300	o
pedal_0097	00:00:00.300	00:00:00.400	o
chap_0001	00:00:00.400	00:00:00.450	o
pedal_0098	00:00:00.450	00:00:00.500	o
pedal_0100	00:00:00.500	00:00:00.600	o
pedal_0101	00:00:00.600	00:00:00.700	o
pedal_0102	00:00:00.700	00:00:00.800	o
pedal_0103	00:00:00.800	00:00:00.900	o

(*Id.*, Fig. 5.)

256. 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 (McCartney) ¶¶[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].)

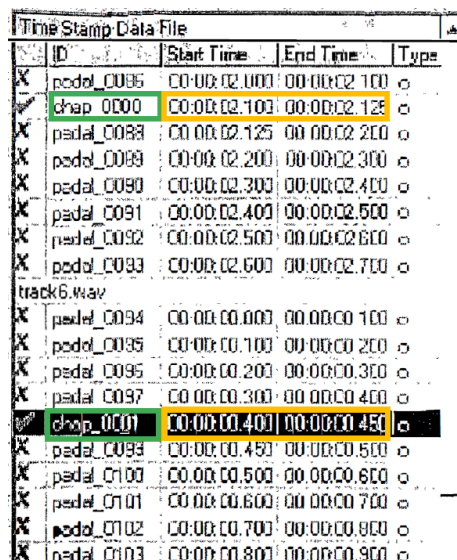
257. Accordingly, McCartney discloses or renders obvious selecting synchronization time offsets (e.g., clip start and end attributes for linked text and audio

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). Thus, McCartney discloses or renders obvious claim element 1[f].

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

258. 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.” McCartney discloses this claim element.

259. 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):



ID	Start Time	End Time	Type
padel_0085	00:00:02.000	00:00:02.100	
chap_0000	00:00:02.100	00:00:02.125	
padel_0088	00:00:02.125	00:00:02.200	
padel_0089	00:00:02.200	00:00:02.300	
padel_0090	00:00:02.300	00:00:02.400	
padel_0091	00:00:02.400	00:00:02.500	
padel_0092	00:00:02.500	00:00:02.600	
padel_0093	00:00:02.600	00:00:02.700	
track6.wav			
padel_0094	00:00:00.000	00:00:00.100	
padel_0095	00:00:00.100	00:00:00.200	
padel_0095	00:00:00.200	00:00:00.300	
padel_0097	00:00:00.300	00:00:00.400	
chap_0001	00:00:00.400	00:00:00.450	
padel_0098	00:00:00.450	00:00:00.500	
padel_0100	00:00:00.500	00:00:00.600	
padel_0101	00:00:00.600	00:00:00.700	
padel_0102	00:00:00.700	00:00:00.800	
padel_0103	00:00:00.800	00:00:00.900	

(EX-1013 (McCartney), Fig. 5; *id.* ¶[0033] (ID attributes correspond to clip start/end times).)

260. Thus, McCartney discloses claim element 1[g][i].

I. Element 1[g][ii]: Synchronized Rendering

261. 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.” McCartney discloses or renders obvious this claim element.

262. 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 (McCartney) ¶¶[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.)

263. 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 (McCartney) ¶¶[0057]-[0058].) 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.*)

264. 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). Thus, McCartney discloses or renders obvious claim element 1[g][ii].

265. Accordingly, for at least the reasons I discuss above, McCartney renders claim 1 obvious as a whole.

2. Claim 2

266. Claim 2 depends from claim 1 and further recites “storing the descriptor file on a server that is accessible to the client device.” McCartney discloses or renders obvious the additional claim element of claim 2.

267. McCartney discloses a system comprising: (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 (McCartney), Fig. 1, ¶¶[0022]-[0025].) 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.” (*Id.* ¶¶[0025] (TOAC general purpose computer may include a server), [0023] (preprocessing unit’s general purpose computer may be “similar to that of the TOAC”).)

268. McCartney’s TSD and BPM files are stored, at least temporarily, on the pre-processing unit because the pre-processing unit creates those files. (EX-1013

(McCartney) ¶¶[0025], [0023].) The files are also stored on, and used by, the TOAC. (*Id.*) It would have been obvious to a person of ordinary skill in the art 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).

269. A person of ordinary skill would have been motivated to store McCartney's TSD and BPM files on a server accessible to TOAC's input/output computer. A person of ordinary skill would have understood that storing the TSD and BPM files on a server allows increased accessibility. For example, multiple TOAC users could access and edit any given TSD/BPM file. Moreover, any given TOAC user could access the TSD/BPM files from any TOAC system. A person of ordinary skill 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).

270. Thus, McCartney discloses or renders obvious the additional claim element of claim 2.

271. Accordingly, for at least the reasons I discuss above, McCartney renders claim 2 obvious as a whole.

3. Claim 3

272. 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.” McCartney discloses or renders obvious the additional claim element of claim 3 for the reasons I discuss above. (*See* ¶¶261-265, above.)

273. Accordingly, for at least the reasons I discuss above, McCartney renders claim 3 obvious as a whole.

4. Claim 8

274. 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.” McCartney discloses or renders obvious the additional claim element of claim 8.

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

276. I understand McCartney incorporates Woodill by reference. (EX-1013 (McCartney) ¶[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 (Woodill) ¶¶[0001], [0003], [0027]-[0029], [0036]-[0038], claim 1.) Time offsets are then extracted and identified as corresponding to pages or chapters. (*Id.*)

277. Accordingly, McCartney discloses (via Woodill), or renders obvious, identifying a plurality of time offsets in dependence on natural language gaps (e.g., silence). Thus, McCartney discloses or renders obvious the additional claim element of claim 8.

278. Accordingly, for at least the reasons I discuss above, McCartney renders claim 8 obvious as a whole.

5. Claims 9-10

279. 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.” 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.” McCartney discloses or renders obvious the additional claim elements of claims 9 and 10.

280. McCartney’s audio data “is typically independently-produced audio clips” or files. (EX-1013 (McCartney) ¶¶[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.)

281. Accordingly, 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). Thus, McCartney discloses or renders obvious the additional claim elements of claims 9 and 10.

282. Accordingly, for at least the reasons I discuss above, McCartney renders claims 9 and 10 obvious as a whole.

6. Claims 11-12

a. Element 11[a]: Selecting Two or More of the Streams

283. Claim 11 depends from claim 1. Element 11[a] recites “selecting two or more of the plurality of digital media streams.” McCartney discloses or renders obvious this claim element.

284. McCartney discloses “[an] operator may select a book section to work with and open it.” (EX-1013 (McCartney) ¶¶[0038]-[0039].) Alternatively, a person of ordinary skill would have understood a user could select a book using a media player. In each case, by selecting a book, the operator or user is selecting two or more of the plurality of streams. Thus, McCartney discloses or renders obvious claim element 11[a].

**b. Element 11[b]: Determining a First Position and
Claim 12: User Interaction**

285. 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.” 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.” McCartney discloses or renders obvious these claim elements.

286. 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 (McCartney) ¶¶[0009] (“operator can click on an event in the TSD list to hear the audio clip represented by that event while the associated HTML element (if any) is highlighted”), [0046], [0042]-[0043], [0024].) 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 id.* ¶¶[0009] (clicking on linked events “caus[es] the TOAC to begin playing the associated audio clip”), [0042].)

287. Thus, McCartney discloses or renders obvious claim elements 11[b] and 12.

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

288. 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.” McCartney discloses or renders obvious this claim element because it discloses that 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 (McCartney) ¶[0046]; *see* ¶¶261-264, above.)

289. Accordingly, for at least the reasons I discuss above, McCartney renders claims 11 and 12 obvious as a whole.

7. Claim 14

290. Claim 14 depends from claim 1 and further recites “wherein the descriptor file contains information that is human readable when rendered.” McCartney discloses or renders obvious the additional claim element of claim 14.

291. 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, as I discuss above. (*See* ¶¶226-231, above.)

292. 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 (McCartney), Fig. 5, 11-13.)

293. 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 for the reasons I discuss above. (*See* ¶¶134-137, 198-206, above; EX-1013 (McCartney), 7, 12.)

294. Thus, McCartney discloses or renders obvious the additional claim element of claim 14.

295. Accordingly, for at least the reasons I discuss above, McCartney renders claim 14 obvious as a whole.

8. Claim 17

296. 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.” McCartney discloses the additional claim element of claim 17 for the reasons I discuss above. (See ¶¶225-265, above.)

297. Accordingly, for at least the reasons I discuss above, McCartney renders claim 17 obvious as a whole.

9. Claim 18

298. Claim 18 depends from claim 1. 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.” 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.” McCartney discloses or renders obvious these claim elements.

299. As I discuss above, 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). (See ¶¶261-265, 283-289, above.)

300. As I further discuss above, storing media content such that it was not resident on the TOAC would have been obvious in view of McCartney. (See ¶¶266-271, above.)

301. In addition, a person of ordinary skill would have understood content was either stored locally or remotely, both of which would have been obvious.

302. Thus, McCartney discloses or renders obvious the additional claim elements of claim 18.

303. Accordingly, for at least the reasons I discuss above, McCartney renders claim 18 obvious as a whole.

10. Claim 19

304. 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.” McCartney discloses or renders obvious the additional claim element of claim 19.

305. McCartney discloses rendering discontinuous audio and text streams (e.g., made up of multiple files). (EX-1013 (McCartney) ¶¶ [0032] (“The TSD file 300 contains one *or more* <data> elements 310, where each data element contains the time data for a single audio file” (emphasis added)), [0023], 7-9 (indicating multiple source text files for a given book (e.g., book project)), Fig. 3.) McCartney never discloses reassembling the separate/discontinuous files into a single file prior to rendering. Thus, McCartney discloses or renders obvious the additional claim element of claim 19.

306. Accordingly, for at least the reasons I discuss above, McCartney renders claim 19 obvious as a whole.

11. Claim 21

307. As I discuss above, independent claim 21 recites substantially the same limitations as claim 1, but 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 method steps of claim 1.

308. McCartney discloses or renders obvious claim 21’s preamble. (EX-1013 (McCartney) ¶¶[0022]-[0025] (computer including a pre-processing unit to generate synchronization files).)

309. The actions recited in claim 21 mirror the method steps recited in claim 1. (See ¶153, above.) As I discuss above for claim 1, McCartney discloses or renders obvious the actions recited in claim 21. (See ¶¶225-265, above.)

310. Accordingly, for at least the reasons I discuss above, McCartney renders claim 21 obvious as a whole.

I. Claims 2 and 18 Would Have Been Obvious in View of McCartney and Kauffman.

311. As I discuss above, McCartney discloses or renders obvious each limitation of claims 2 and 18. (See ¶¶266-271, 298-303, above.) For at least those reasons and the reasons I discuss below, claims 2 and 18 would also have been obvious

in further view of Kauffman. I incorporate my above discussion of McCartney here, and only discuss claim elements for which Kauffman is relevant in combination with McCartney.

1. Claim 2

312. Claim 2 depends from claim 1 and further recites “storing the descriptor file on a server that is accessible to the client device.” Kauffman discloses the additional claim element of claim 2.

313. Kauffman discloses an object server storing a media object divided into a plurality of pieces and a “piece map.” (EX-1015 (Kauffman), 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 (object server sends a copy of the piece map object to library client).) Accordingly, 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). Thus, Kauffman discloses the additional claim element of claim 2.

314. A person of ordinary skill in the art would have been motivated to store McCartney’s TSD and BPM files on a server accessible to TOAC’s input/output computer for multiple reasons.

315. First, a person of ordinary skill would have been motivated to do so for the reasons I discuss above. (*See* ¶¶266-271, above.)

316. Second, 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). Third, 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. Fourth, 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).

317. Finally, storing descriptor files on servers was commonplace and disclosed in other references, giving a person of ordinary skill a reasonable expectation of success in the modification. (*See, e.g.*, EX-1106 (Chen), 889-91; EX-1006 (Yoshimura).)

318. Accordingly, for at least the reasons I discuss above, McCartney and Kauffman render claim 2 obvious as a whole.

2. Claim 18

319. Claim 18 depends from claim 1. 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.” 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.” Kauffman discloses or renders obvious these claim elements for the reasons I discuss above. (*See* ¶¶312-318, above.)

320. A person of ordinary skill in the art would have been motivated to store content remotely for the reasons discussed above (*see* ¶¶312-318, above) 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). A person of ordinary skill would have reasonably expected success in making the modification because storing and transferring media streams from remote servers was well known and conventional. (*See, e.g.*, EX-1029 (Steele), 2-3, Fig. 2; EX-1106 (Chen), 888-89.)

321. Thus, for at least the reasons I discuss above, McCartney and Kauffman render claim 18 obvious as a whole.

J. Claims 4-6 and 13 Would Have Been Obvious in View of McCartney and DTB.

322. Claims 4-6 and 13 depend, directly or indirectly, from claims 3 or 11. As I discuss above, McCartney discloses or renders obvious each limitation of claims 3 and 11. (*See* ¶¶272-273, 283-289, above.) DTB discloses the additional claim elements of claims 4-6 and 13.

1. Claim 4

a. Element 4[a]: Tracking Current Position

323. 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.” McCartney and DTB each disclose or render obvious this claim element.

324. 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 (McCartney) ¶¶[0037]-[0056].) As discussed above, McCartney’s system pre-populates links between sections (e.g., chapters) of the audio and text streams. (*Id.* ¶¶[0041]-[0044].) McCartney then asks a user whether they want to check the accuracy of each link. (*Id.* ¶[0045].) 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. (*Id.* ¶¶[0044]-[0046], [0056]-[0058], Fig. 6.) McCartney therefore discloses or renders obvious claim element 4[a].

325. DTB also discloses or renders obvious claim element 4[a] for the reasons I discuss above. (*See* ¶¶94-98, above.) A person of ordinary skill in the art would have been motivated to track a current position in McCartney’s media streams for multiple reasons.

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

327. 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). 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. 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).

328. A person of ordinary skill 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. (*See, e.g.*, EX-1033 (Hanson), 5:4-6 ("the system monitors the time ... that the user has progressed into the audio").)

b. Element 4[b]: Bookmark

329. 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 discloses or renders obvious this claim element for the reasons I discuss above. (*See* ¶¶94-98, above.)

330. A person of ordinary skill in the art 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. For example, McCartney’s TOAC “can operate on one ‘book section’ at a time.” (EX-1013 (McCartney) ¶[0039].) If a user wishes to move to another section of the book, or a different book altogether, a person of ordinary skill 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).

331. A person of ordinary skill would have been motivated to create a bookmark and would have had a reasonable expectation of success in doing so in McCartney’s system for the reasons I discuss above. (*See* ¶¶323-328, above.)

332. Accordingly, for at least the reasons I discuss above, McCartney and DTB render claim 4 obvious as a whole.

2. Claim 5

333. 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 discloses or renders obvious the additional claim element of claim 5.

334. As I discuss above, DTB discloses rendering one or more of the plurality of digital media streams starting from the bookmarked position on the original rendering device or a second other client device. (See ¶¶99-101, above.) Thus, DTB discloses or renders obvious the additional claim element of claim 5.

335. It would have been obvious to a person of ordinary skill to use the bookmark I discuss above for claim 4 to render the content on the same or a different TOAC device. (See ¶¶323-332, above.)

336. Accordingly, for at least the reasons I discuss above, McCartney and DTB render claim 5 obvious as a whole.

3. Claim 6

337. 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.” As I discuss above, DTB discloses or renders obvious the additional claim element of claim 6. (See ¶¶102-104, above.) As in DTB, McCartney’s bookmark could be based on the text stream *or* the audio stream, not necessarily both.

338. A person of ordinary skill would have been motivated to combine McCartney and DTB and would have had a reasonable expectation of success in doing so for the reasons I discuss above. (See ¶¶323-332, above.)

339. Accordingly, for at least the reasons I discuss above, McCartney and DTB render claim 6 obvious as a whole.

4. Claim 13

340. 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 or renders obvious the additional claim element of claim 13 for the reasons I discuss above. (See ¶¶132-133, above.)

341. A person of ordinary skill in the art would have understood that the DTB/McCartney bookmark I discuss above would include either the ID of the TSD element or the absolute time being marked as well as an identification of the TSD file. (See ¶¶329-331, above.) In either case, the position would be determined 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.

342. A person of ordinary skill would have been motivated to combine McCartney and DTB and would have had a reasonable expectation of success in doing so for the reasons I discuss above. (See ¶¶323-336, 258-260 (TSD file stores synchronization time offsets).)

343. Accordingly, for at least the reasons I discuss above, McCartney and DTB render claim 13 obvious as a whole.

K. Claim 7 Would Have Been Obvious in View of McCartney and Duncan.

344. Claim 7 depends from claim 3. As I discuss above, McCartney discloses or renders obvious each limitation of claim 3. (See ¶¶272-273, above.) McCartney and/or Duncan disclose the additional claim elements of claim 7.

345. 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.” McCartney discloses this claim element because McCartney discloses visually rendered content (e.g., text). (EX-1013 (McCartney) ¶¶[0044]-[0046].)

346. Element 7[b] recites “providing automated page turns as the two or more of the plurality of digital media streams are simultaneously rendered.” Duncan discloses this claim element for the reasons I discuss above. (See ¶¶179-187, above.) It would have been obvious to a person of ordinary skill in the art to use Duncan’s auto page turn feature on McCartney’s review at the TOAC computer. (*Id.*)

347. A person of ordinary skill in the art would have been motivated to combine Duncan’s automatic page turning with McCartney’s system for several reasons.

348. First, McCartney discloses a method for producing digital talking books, and Duncan discloses a method for presenting them. (EX-1013 (McCartney), Abstract; EX-1027 (Duncan), Abstract.)

349. 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 (Duncan) ¶[0042]; EX-1013 (McCartney) ¶[0009] (TOAC operator can verify that text and audio have been correctly linked).)

350. 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). 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. 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).

351. A person of ordinary skill in the art 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 (Duncan) ¶¶[0041]-[0042]), and McCartney discloses identifying page events in an XML text stream. (EX-1013 (McCartney) ¶[0041] (method of McCartney correlates

“page elements” in the text with “page events” in the audio); *id.* ¶[0029] (text can be in XML format).)

352. Accordingly, for at least the reasons I discuss above, McCartney and Duncan render claim 7 obvious as a whole.

L. Claim 8 Would Have Been Obvious in View of McCartney and Woodill.

353. As I discuss above, McCartney discloses or renders obvious each limitation of claim 8. (*See* ¶¶274-278, above.) For example, as I discuss above, McCartney discloses (via Woodill), or renders obvious, identifying a plurality of time offsets in dependence on natural language gaps (e.g., silence). (*Id.*) For at least those reasons and the reasons I discuss below, claim 8 would also have been obvious in further view of Woodill. I incorporate my above discussion of McCartney here.

354. 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.” Woodill discloses the additional claim element of claim 8 for the reasons I discuss above. (*See* ¶¶274-278, above.)

355. A person of ordinary skill in the art would have been motivated to combine McCartney and Woodill for several reasons.

356. 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 (McCartney) ¶[0022].) Second, both references are directed to creating TSD files. (EX-1014 (Woodill) ¶[0024]; EX-1013 (McCartney), Abstract.) Third, McCartney and Woodill share inventors. (EX-1013 (McCartney), 1; EX-1014 (Woodill), 1.) A person of ordinary skill would have had a reasonable expectation of success in combining McCartney and Woodill for the same reasons.

357. Accordingly, for at least the reasons I discuss above, McCartney and Woodill render claim 8 obvious as a whole.

M. Claims 14 and 20 Would Have Been Obvious in View of McCartney and Bulterman.

358. As I discuss above, McCartney discloses or renders obvious each limitation of claim 14. (See ¶¶290-295, above.) For at least those reasons and the reasons I discuss below, claim 14 would also have been obvious in further view of Bulterman. I incorporate my above discussion of McCartney here. Furthermore, claim 20 depends from claim 1. As I discuss above, McCartney discloses or renders obvious each limitation of claim 1. (See ¶¶225-265, above.) For at least those reasons and the reasons I discuss below, claim 20 would have been obvious in further view of Bulterman.

1. Claim 14

359. Claim 14 further recites “wherein the descriptor file contains information that is human readable when rendered.” Bulterman discloses the additional claim element of claim 14 for the reasons I discuss above. (See ¶¶198-206, above.)

360. A person of ordinary skill in the art would have been motivated to combine Bulterman’s inclusion of the actual text (as opposed to a link, path, or ID) in McCartney’s TSD and/or BPM file for the reasons I discuss above. (See ¶¶198-206, above; EX-1101 (Simpson), 5.) A person of ordinary skill 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 (Simpson), 2-5; EX-1013 (McCartney), 7, 12.) 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. (See EX-1013 (McCartney) ¶[0057] (SMIL files can be generated from TSD files); *id.* ¶[0030], 7 (BPM includes “<smil_sync>” element); EX-1003 (DTB), 23 (DTB SMIL files are “XML documents”).)

361. Accordingly, for at least the reasons I discuss above, McCartney and Bulterman render claim 14 obvious as a whole.

2. Claim 20

362. 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.” Bulterman discloses the additional claim element of claim 20 for the reasons I discuss above. (*See* ¶¶207-209, above.)

363. Including the actual text in McCartney’s BPM “path” field (e.g., McCartney’s location information) would have been obvious to a person of ordinary skill in the art. (*See* ¶¶198-206, above.) The presence of text (as opposed to a file path) would indicate to a person of ordinary skill that the text stream is located in the BPM file. Alternatively, it would have been obvious to a person of ordinary skill to include an indicator like that disclosed in Bulterman. A person of ordinary skill 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, for the reasons I discuss above. (*See* ¶¶198-206, 359-361, above.)

364. Accordingly, for at least the reasons I discuss above, McCartney and Bulterman render claim 20 obvious as a whole.

N. Claim 15 Would Have Been Obvious in View of McCartney and Shteyn.

365. 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.” As I discuss above, McCartney discloses or renders obvious

each limitation of claim 1. (*See* ¶¶225-265, above.) Shteyn discloses the additional claim element of claim 15.

366. 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 (Shteyn), Abstract, 3:1-2, 3:27-4:25, claim 1.) Thus, Shteyn discloses the additional claim element of claim 15.

367. It would have been obvious to a person of ordinary skill in the art 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. (EX-1086 (Shteyn), Abstract, 3:1-2, 3:27-4:25, claim 1.) I discuss the benefits of doing so above. (*See* ¶¶210-216, above.) 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 (Shteyn), 2:11-3:7.) A person of ordinary skill 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., id.*, Abstract,

1:13-14, 2:20-22 (explaining that “splitter[s]” and clients for processing split data were available).)

368. Accordingly, for at least the reasons I discuss above, McCartney and Shteyn render claim 15 obvious as a whole.

O. Claim 16 Would Have Been Obvious in View of McCartney, Shteyn, and Copley.

369. 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.” As I discuss above, McCartney and Shteyn disclose or render obvious each limitation of claim 15. (See ¶¶365-368, above.) Copley discloses the additional claim element of claim 16, as I also discuss above. (See ¶¶217-223, above.)

370. It would have been obvious to a person of ordinary skill in the art to manage retrieval of the data in McCartney by Copley’s method. (See ¶¶217-223, above.) In addition to the reasons I discuss above, a person of ordinary skill 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 (McCartney) ¶[0046]; EX-1009 (Copley) ¶¶[0013]-[0015].)

371. Accordingly, for at least the reasons I discuss above, McCartney, Shteyn, and Copley render claim 16 obvious as a whole.

VII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

372. I am not aware of any secondary considerations of nonobviousness. If Patent Owner identifies any alleged evidence of secondary considerations in the future, I reserve the right to respond to that information.

VIII. CONCLUSION

373. For the foregoing reasons, it is my opinion that claims 1-21 of the '907 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of the prior art discussed above.

374. I reserve the right to supplement my opinions in the future to address or respond to any issues that the Patent Owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the Patent Owner or adopted by the Board in the Institution Decision, and respond to any alleged secondary considerations as they become available to me.

Amazon.com, Inc. v. Audio Pod IP, LLC
Declaration of Ketan Mayer-Patel – U.S. Pat. No. 9,729,907

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on May 8th, 2025

at Chapel Hill, NC.

A handwritten signature in black ink, appearing to read 'JKMP', written over a horizontal line.

Professor Ketan Mayer-Patel, Ph.D.