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. IPR2025-00777
U.S. Patent No. 9,319,720

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,319,720

TABLE OF CONTENTS

I. INTRODUCTION ---------- 1

II. BACKGROUND ---------- 1

A. Segmenting Media Content Was Known ---------- 1

B. Descriptor Files for Synchronizing Media Content Were Known ---------- 2

C. Identifying a File Using a Time Offset and a Descriptor File Was Known ---------- 3

III. THE '720 PATENT ---------- 3

A. Overview ---------- 3

B. Prosecution ---------- 4

C. Priority ---------- 4

IV. LEVEL OF ORDINARY SKILL IN THE ART ---------- 4

V. CLAIM CONSTRUCTION ---------- 5

VI. STATEMENT OF PRECISE RELIEF REQUESTED ---------- 5

A. Grounds ---------- 5

B. Status of References as Prior Art ---------- 6

VII. GROUND 1A: CLAIMS 1, 18, AND 25-26 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA AND LINDAHL ---------- 7

A. Claim 1 ---------- 7

1. Preamble ---------- 7

2. Element-1[a] ---------- 7

3. Element-1[b] ----------8
a. Descriptor File ----------8
b. Media Streams ---------- 10
c. Same Originating Work---------- 11
i. Yoshimura---------- 11
ii. Lindahl---------- 11
iii. Motivation to Combine---------- 12
4. Element-1[c] ---------- 13
5. Element-1[d] ---------- 14
6. Element-1[e] ---------- 15
7. Element-1[f]---------- 17
a. Yoshimura ---------- 17
b. Lindahl ---------- 18
c. Motivation to Combine ---------- 20
8. Element-1[g] ---------- 21
9. Element-1[h] ---------- 22
B. Claim 18---------- 22
C. Claim 25---------- 23
D. Claim 26---------- 23
VIII. GROUND 1B: CLAIMS 2-10, 14-15, AND 22 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND HECKERMAN---------- 23

A. Claim 2 ---------- 23

B. Claim 3 ---------- 25

C. Claim 4 ---------- 26

D. Claim 5 ---------- 27

E. Claim 6 ---------- 28

1. Element-6[a] ---------- 28

2. Element-6[b] ---------- 28

F. Claim 7 ---------- 29

G. Claim 8 ---------- 29

H. Claim 9 ---------- 30

I. Claim 10 ---------- 30

J. Claim 14 ---------- 30

K. Claim 15 ---------- 31

L. Claim 22 ---------- 31

IX. GROUND 1C: CLAIMS 5 AND 11-13 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND BULTERMAN ---------- 31

A. Claim 5 ---------- 31

B. Claim 11 ---------- 32

1. Element-11[a] ---------- 32

2. Element-11[b] ---------- 32

C. Claim 12 ---------- 35

D. Claim 13 ---------- 38

X. GROUND 1D: CLAIM 17 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND DTB ---------- 38

XI. GROUND 1E: CLAIM 20 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND DTB ---------- 40

XII. GROUND 1F: CLAIMS 16 AND 24 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND HAY ---------- 41

A. Claim 16 ---------- 41

B. Claim 24 ---------- 42

XIII. GROUND 1G: CLAIMS 19, 23, AND 27 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND SULL-948 ---------- 42

A. Claim 19 ---------- 42

B. Claim 23 ---------- 44

C. Claim 27 ---------- 45

XIV. GROUND 1H: CLAIM 21 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND AMIR---------- 45

XV. GROUND 1I: CLAIMS 28 AND 29 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, SULL-948, AND DTB ---------- 46

A. Claim 28 ---------- 46

B. Claim 29 ---------- 48

XVI. GROUND 2A: CLAIMS 1-15, 17-18, 20-22, AND 25-26 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, AND OPTIONALLY STELOVSKY ---------- 49

A. Claim 1 ---------- 49

1. Preamble ---------- 49

2. Element-1[a] ---------- 49

3. Element-1[b] ---------- 49

4. Element-1[c] ---------- 52

5. Element-1[d] ---------- 53

6. Element-1[e] ---------- 53

7. Element-1[f] ---------- 57

8. Element-1[g] ---------- 58

9. Element-1[h] ---------- 59

B. Claim 2 ---------- 60

C. Claim 3 ---------- 60

1. Audio Index ---------- 60

2. Shared Audio/Slide Index ---------- 62

D. Claim 4 ---------- 63

E. Claim 5 ---------- 63

1. Element-5[a] ---------- 63

2. Element-5[b] ---------- 64

F. Claim 6 ---------- 64

1. Element-6[a] ---------- 64

2. Element-6[b] ---------- 64

G. Claim 11 ---------- 65

1. Element-11[a] ---------- 65

2. Element-11[b] ---------- 65

H. Claim 12 ---------- 65

I. Claim 13 ---------- 66

J. Claims 14 and 18 ---------- 67

K. Claim 17 ---------- 68

L. Claim 22 ---------- 68

M. Claims 7-10, 15, 20-21, and 25-26 ---------- 69

XVII. GROUND 2B: CLAIMS 16 AND 24 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, HAY, AND OPTIONALLY STELOVSKY ---------- 69

XVIII. GROUND 2C: CLAIMS 19, 23, AND 27 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, SULL-948, AND OPTIONALLY STELOVSKY ---------- 70

XIX. GROUND 2D: CLAIMS 28 AND 29 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, SULL-948, DTB, AND OPTIONALLY STELOVSKY ---------- 70

XX. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS ---------- 71

XXI. DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE ---------- 71

A. Factor 1: Potential Stay ---------- 71

B. Factor 2: Proximity of Trial to FWD ---------- 72

C. Factor 3: Investment in Parallel Proceeding ---------- 73

D. Factor 4: Overlapping Issues ---------- 74

E. Factor 5: The Parties ---------- 75

F. Factor 6: Other Circumstances ---------- 75

XXII. DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE ---------- 75

XXIII. CONCLUSION ---------- 76

XXIV. MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT ---------- 76

A. Real Parties-In-Interest (37 C.F.R. §42.8(b)(1)) ---------- 76

B. Related Matters (37 C.F.R. §42.8(b)(2)) ---------- 76

C. Lead and Backup Counsel (37 C.F.R. §42.8(b)(3))---------- 77

D. Service Information (37 C.F.R. §42.8(b)(4)) ---------- 78

E. Grounds for Standing (37 C.F.R. §42.104(a)) ---------- 78

F. Payment of Fees (37 C.F.R. §42.15(a)) ---------- 79

TABLE OF AUTHORITIES

Cases: *Page(s):*

Amazon.com, Inc. v. Nokia Technologies OY,
IPR2024-01140, Paper 9 (P.T.A.B. Feb. 12, 2025) ---------- 72, 74

Apple Inc. v. Fintiv, Inc.,
IPR2020-00019, Paper 11 ---------- 71, 75

Aptiv Services US, LLC v. Microchip Technology, Inc.,
IPR2024-00646, Paper 11 (P.T.A.B. Sept. 25, 2024) ---------- 73

Cal. Inst. of Tech. v. Broadcom Ltd.,
25 F.4th 976 (Fed. Cir. 2022) ---------- 74

Ericsson Inc. v. XR Communications LLC,
IPR2024-00613, Paper 9 (P.T.A.B. Oct. 9, 2024) ---------- 73, 74

Google LLC v. Jawbone Innovations, LLC,
IPR2022-00630, Paper 10 (P.T.A.B. Sept. 13, 2022) ---------- 75

In re GPAC Inc.,
57 F.3d 1573 (Fed. Cir. 1995) ---------- 5

KSR Int'l Co. v. Teleflex Inc.,
550 U.S. 398 (2007) ----------*passim*

Leapfrog Enters., Inc. v. Fisher-Price, Inc.,
485 F.3d 1157 (Fed. Cir. 2007) ---------- 71

Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.,
868 F.3d 1013 (Fed. Cir. 2017) ---------- 5

Quasar Sci. LLC v. Colt Int'l Clothing, Inc.,
IPR2023-00611, Paper 10 (P.T.A.B. Oct. 10, 2023) ---------- 76

Samsung Electronics Co. v. Empire Technology Development LLC,
IPR2024-00896, Paper 15 (P.T.A.B. Dec. 13, 2024) ---------- 74

Sec. First Innovations, LLC v. Google LLC,
No. 2:23-cv-00097, 2024 WL 234720 (E.D. Va. Jan. 22, 2024) ---------- 71

Sharpe Innovations, Inc. v. T-Mobile USA, Inc.,
No. 2:17-cv-00351, 2018 WL 11198604 (E.D. Va. Jan. 10, 2018) ---------- 72

Shenzen Chic Elecs. v. Pilot, Inc.,
IPR2023-00810, Paper 12 (P.T.A.B. Nov. 8, 2023)---------- 75

Sotera Wireless, Inc. v. Masimo Corp.,
IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020)---------- 74

Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.,
200 F.3d 795 (Fed. Cir. 1999)---------- 5

Statutes and Rules:

35 U.S.C. §102---------- 6

35 U.S.C. §103---------- 5

35 U.S.C. §314---------- 71, 75

35 U.S.C. §325---------- 75, 76

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 9,319,720 ("the '720 patent")
1002	Declaration of Professor Ketan Mayer-Patel, Ph.D.
1003	NATIONAL INFORMATION STANDARDS ORGANIZATION, SPECIFICATIONS FOR THE DIGITAL TALKING BOOK (ANSI/NISO Z39.86-2002) (2002) ("DTB")
1004	European Patent Publication No. EP 1463258 A1 ("Lindahl")
1005	U.S. Patent Publication No. 2002/0069218 ("Sull-218")
1006	Yoshimura et al., *Content Delivery Network Architecture for Mobile Streaming Service Enabled by SMIL Modification*, 86 IEICE TRANSACTIONS ON COMMC'N 1778 (2003) ("Yoshimura")
Exhibit Numbers 1007-1011 Not Used	
1012	Excerpts from DICK C.A. BULTERMAN & LLOYD RUTLEDGE, SMIL 2.0, INTERACTIVE MULTIMEDIA FOR WEB AND MOBILE DEVICES (2004) ("Bulterman")
Exhibit Numbers 1013-1027 Not Used	
1028	U.S. Patent Publication No. 2002/0184189 ("Hay")
Exhibit Number 1029 Not Used	
1030	U.S. Patent No. 6,260,011 ("Heckerman")
Exhibit Numbers 1031-1039 Not Used	

Exhibit No.	Description
1040	Amir et al., *Search the Audio, Browse the Video—A Generic Paradigm for Video Collections*, 2003 EURASIP J. ON APPLIED SIGNAL PROCESSING 209 (2003) ("Amir")
Exhibit Numbers 1041-1042 Not Used	
1043	U.S. Patent No. 5,535,063 ("Lamming")
1044	U.S. Patent No. 5,613,909 ("Stelovsky")
Exhibit Number 1045 Not Used	
1046	Tredinnick, *Implementing the SMIL Specification*, PROC. OF THE LINUX SYMP. (2003) ("Tredinnick")
1047	U.S. Patent Publication No. 2007/0150930 ("Koivisto")
1048	U.S. Patent No. 7,412,643 ("Fischer")
Exhibit Numbers 1049-1050 Not Used	
1051	Excerpts from World Wide Web Consortium, *Synchronized Multimedia Integration Language (SMIL 2.0) Specification*, W3C Recommendation (Aug. 7, 2001), https://www.w3.org/TR/2001/REC-smil20-20010807/smil20.html. (the "SMIL Standard")
Exhibit Numbers 1052-1056 Not Used	
1057	U.S. Patent No. 5,798,841 ("Takahashi")
1058	U.S. Patent No. 5,021,989 ("Fujisawa")
1059	PCT Patent Publication No. WO2001/01373 ("Hendricks")

Exhibit No.	Description
1060	PCT Patent Publication No. WO2002/08948A2 ("Sull-948")
Exhibit Number 1061 Not Used	
1062	U.S. Patent No. 6,208,995 ("Himmel")
1063	P. Delacourt & C.J. Wellekens, *DISTBIC: A speaker-based segmentation for audio data indexing*, 32 SPEECH COMMUNICATION 111 (2000) ("Delacourt")
Exhibit Numbers 1064-1068 Not Used	
1069	PCT Patent Publication No. WO2002/080524A2 ("Dimitrova")
Exhibit Numbers 1070-1075 Not Used	
1076	PCT Patent Publication No. WO1997/41504A1 ("Reid")
Exhibit Numbers 1077-1079 Not Used	
1080	U.S. Patent Publication No. 2003/0091338 ("Snow")
Exhibit Numbers 1081-1084 Not Used	
1085	U.S. Patent Publication No. 2002/0147979 ("Corson")
1086	PCT Patent Publication No. WO2001/24474 ("Shteyn")
Exhibit Numbers 1087-1090 Not Used	
1091	C. Duarte & Luís Carriço, *Usability Evaluation of Digital Talking Books*, PROC. 1ST NAT'L CONF. ON HUMAN-COMPUTER INTERACTION 2004 (June 2004) ("Duarte")

Exhibit No.	Description
Exhibit Numbers 1092-1094 Not Used	
1095	Excerpts from the File History for U.S. Patent No. 9,319,720
1096	CV of Professor Ketan Mayer-Patel, Ph.D.
1097	Declaration of Sylvia D. Hall-Ellis, Ph.D.
1098	Audio Pod's Infringement Contentions for U.S. Patent No. 9,319,720 in *Audio Pod IP, LLC v. Amazon.com, Inc.*, 3:24-cv-00407 (E.D. Va.)

Petitioners Amazon.com, Inc., Amazon.com Services LLC, Amazon Web Services, Inc., and Audible, Inc. ("Petitioners" or "Amazon") request *inter partes* review of claims 1-29 of U.S. Patent No. 9,319,720 ("the '720 patent"), which Audio Pod IP, LLC ("Patent Owner" or "PO") purportedly owns.

I. INTRODUCTION

The '720 patent relates to rendering digital content on a media player. The claims recite three basic concepts: (1) segmenting media content into files; (2) using a descriptor file with timing information to synchronize media content; and (3) using a time offset and a descriptor file to identify a particular file. These concepts were well-known and their combination would have been obvious. Thus, the Board should cancel the claims.

II. BACKGROUND

A. Segmenting Media Content Was Known.

The '720 patent acknowledges dividing media into separate files was known before the patent's earliest possible priority date in December 13, 2005. (EX-1001, 2:18-23.) Indeed, it was routine. (*See, e.g.*, EX-1006, 1778-81; EX-1003, 6, 17, 19-20; EX-1085, Abstract, ¶¶[0052]-[0054]; EX-1086, Abstract, 2:11-27; EX-1002 ¶32.) For example, Yoshimura disclosed dividing content into segments that are

stored as separate files on servers. (EX-1006, 1779-80, Fig. 3; EX-1002 ¶32.) Heckerman similarly disclosed segmenting audio and text data into multiple audio and text data files. (EX-1030, claim 22; EX-1002 ¶32.)

B. Descriptor Files for Synchronizing Media Content Were Known.

It was also common by 2005 to use descriptor files with timing information to synchronize the playback of media files. (*See generally* EX-1051; EX-1006; EX-1003; EX-1012; EX-1076; EX-1002 ¶33.) In the 1990s, Reid disclosed using descriptor files for synchronizing multimedia presentations. (EX-1076; EX-1002 ¶33.) In 2001, the Synchronized Multimedia Integration Language (SMIL) file, a well-known descriptor file for synchronizing the playback of media files (e.g., text, images or audio) over the Internet, was adopted as a standard. (EX-1051; EX-1012; EX-1002 ¶33.) The below diagram demonstrates how various media streams (e.g., audio, images, etc.) are organized, via a SMIL file, to provide a synchronized presentation, with each media stream's rendering based on the timeline of an audio file:

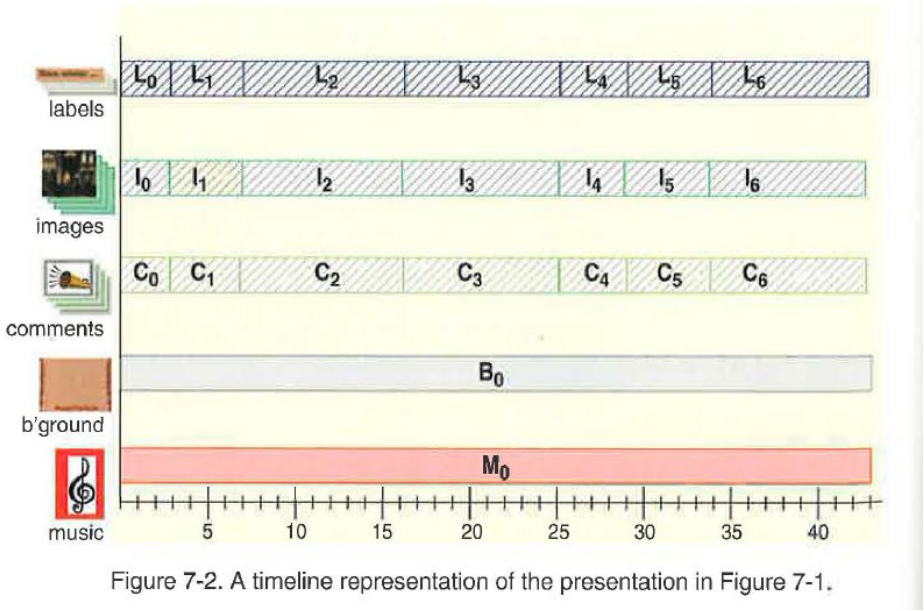


Figure 7-2. A timeline representation of the presentation in Figure 7-1.

(EX-1012, 128; EX-1002 ¶33.)

C. Identifying a File Using a Time Offset and a Descriptor File Was Known.

Digital bookmarks, which identify a particular time (or "time offset") in a media stream, were well known by 2005. (EX-1002 ¶34.) Bookmarks were routinely used with a descriptor file to identify the associated media file for playback or download. (EX-1004 ¶¶[0035]-[0036], [0040]-[0041]; EX-1003, 45-50; EX-1005 ¶¶[0008], [0175]; EX-1047 ¶¶[0003]-[0032]; EX-1002 ¶34.)

III. THE '720 PATENT

A. Overview

The '720 patent describes storing on servers multiple media streams, where each stream is segmented into multiple files. (EX-1001, Abstract, 5:59-6:7; EX-1002 ¶35.) The server also stores a descriptor file with time information about

each file that is used to synchronize the media files. (EX-1001, 6:20-65; EX-1002 ¶35.) The descriptor file identifies particular files for playback using the descriptor file's time information and a time offset provided by the user (e.g., from a book-mark). (EX-1001, 19:51-20:26; EX-1002 ¶35.) The media files can then be down-loaded and rendered synchronously on the media player. (EX-1001, Abstract, 38:13-33; EX-1002 ¶35.)

B. Prosecution

During prosecution, the Examiner repeatedly rejected the pending claims as obvious. (EX-1095, 188-205, 133-56, 63-97.) PO amended the independent claims to add limitations 1[c], 1[d], and 1[e]. (*Id.*, 42-61.) The Examiner then allowed the claims. (*Id.*, 18-24.)

C. Priority

The '720 patent's earliest possible priority date is December 13, 2005. (EX-1001, 1.) Petitioners do not concede the claims are entitled to that priority date.

IV. LEVEL OF ORDINARY SKILL IN THE ART

A POSITA is "a person of ordinary creativity, not an automaton." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 421 (2007). A POSITA would have had at least a bachelor's degree in electrical engineering, computer engineering, or computer science, and at least three years of industry or academic experience in the design,

development, and/or implementation of content rendering and/or distribution systems. (EX-1002 ¶¶27-31); *see In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995). Work experience could substitute for formal education and additional formal education could substitute for work experience. (EX-1002 ¶29.)

V. CLAIM CONSTRUCTION

No claim terms require construction to resolve the obviousness challenges here. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999). For this proceeding only, Petitioner assumes the claims are not invalid under §112 and interprets the claims consistent with PO's Infringement Contentions. (EX-1098.)

VI. STATEMENT OF PRECISE RELIEF REQUESTED

A. Grounds

The Board should cancel the claims under §103 on the following Grounds:

Ground	Claims Challenged	References
1A	1, 18, 25-26	Yoshimura and Lindahl
1B	2-10, 14-15, 22	Yoshimura, Lindahl, and Heckerman
1C	5, 11-13	Yoshimura, Lindahl, Heckerman, and Bulterman
1D	17	Yoshimura, Lindahl, and DTB
1E	20	Yoshimura, Lindahl, Heckerman, and DTB

Ground	Claims Challenged	References
1F	16, 24	Yoshimura, Lindahl, Heckerman, and Hay
1G	19, 23, 27	Yoshimura, Lindahl, and Sull-948
1H	21	Yoshimura, Lindahl, Heckerman, and Amir
1I	28, 29	Yoshimura, Lindahl, Sull-948, and DTB
2A	1-15, 17-18, 20-22, 25-26	Reid, Heckerman, and optionally Stelovsky
2B	16, 24	Reid, Heckerman, Hay, and optionally Stelovsky
2C	19, 23, 27	Reid, Heckerman, Sull-948, and optionally Stelovsky
2C	28-29	Reid, Heckerman, Sull-948, DTB, and optionally Stelovsky

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

B. Status of References as Prior Art

Each reference is prior art under pre-AIA §102(b) because it published more than one year before the '720 patent's earliest possible priority date of December 13, 2005:

Reference	Publication Date	Exhibit
Yoshimura	September 29, 2003	EX-1006; *see* EX-1097
DTB	March 19, 2002	EX-1003; *see* EX-1097
Lindahl	September 29, 2004	EX-1004
Heckerman	July 10, 2001	EX-1030
Bulterman	May 1, 2004	EX-1012; *see* EX-1097

Reference	Publication Date	Exhibit
Stelovsky	March 25, 1997	EX-1044
Amir	February 25, 2003	EX-1040; *see* EX-1097
Hay	December 5, 2002	EX-1028
Sull-948	January 31, 2002	EX-1060
Reid	November 6, 1997	EX-1076

These references are analogous art because they are from the same field of endeavor as the '720 patent, e.g., content distribution and/or rendering. (EX-1002 ¶22. They are also pertinent to a particular problem the inventor was focused on, e.g., efficient and effective distribution and/or rendering of content. (*Id.*)

VII. GROUND 1A: CLAIMS 1, 18, AND 25-26 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA AND LINDAHL.

A. Claim 1

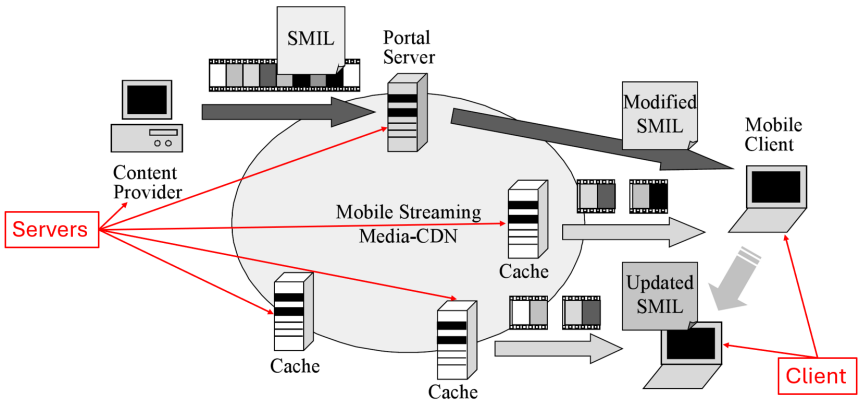
1. Preamble

The preamble recites "[a] method of rendering digital content." Yoshimura discloses "a CDN (Content Delivery Network) for mobile streaming" that delivers "multimedia content" to "mobile clients" to render. (EX-1006, 1778; *id.*, 1780-81.) If the preamble is limiting, Yoshimura discloses it. (EX-1002 ¶¶38-39.)

2. Element-1[a]

Element 1[a] recites "providing a media player having access to at least one server via a network."

Yoshimura discloses a "mobile streaming media CDN" in which clients access servers via a network:



(EX-1006, Fig. 2[1], 1778 (CDN includes servers and client), 1779, 1780 (CDN includes portal servers, content servers, and surrogate servers).) The clients are media players. (*Id.*, 1780, 1781 (content "played and displayed"); EX-1002 ¶41.) Accordingly, Yoshimura discloses providing a media player (e.g., client) having access to at least one server (e.g., content provider, portal, and/or cache/surrogate servers) via a network. (EX-1002 ¶¶40-42.)

3. Element-1[b]

Element 1[b] recites "the at least one server having stored thereon a descriptor file and a plurality of media streams derived from a same originating written work."

a. Descriptor File

Yoshimura's portal server takes a SMIL file attached to "original" content, segments the content into smaller files, and creates a modified SMIL file "to describe the timing relations among these segments" as shown in Figure 3:

[1] Figures herein may be colored or annotated.

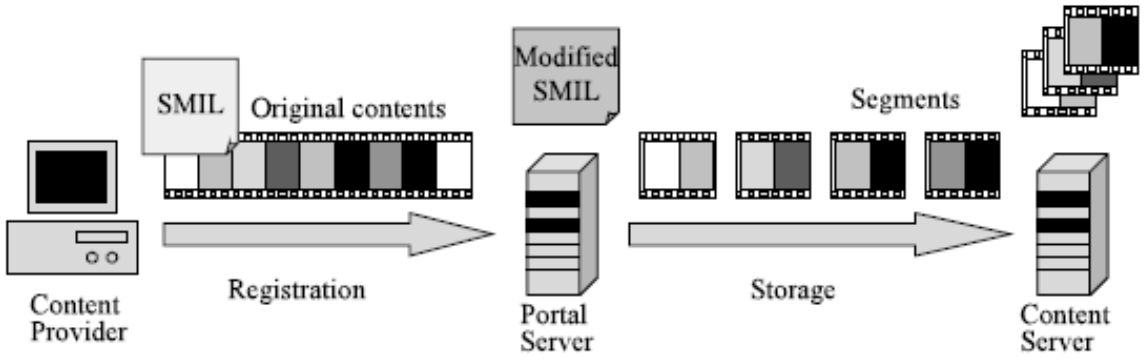


Fig. 3 Content segmentation and SMIL modification at portal server.

(EX-1006, 1780, Fig. 3 ("Modified SMIL" file on portal server); EX-1002 ¶44.) An example of a "modified SMIL" file is shown in Figure 5:

```
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
"http://www.w3.org/2001/SMIL20/SMIL20.dtd">
<smil xmlns="http://www.w3.org/2001/SMIL20/Language">
 <head>
  <layout>
   <region id="v" top="5" left="5" width="180" height="180"/>
  </layout>
 </head>
 <body>
  <seq>
   <par dur="10min">
    <audio src="rtsp://content-server/content-A-1.mp4" />
    <video src="rtsp://content-server/content-V-1.mp4" region="v" />
   </par>
   <par dur="20min">
    <audio src="rtsp://content-server/content-A-2.mp4" />
    <video src="rtsp://content-server/content-V-2.mp4" region="v" />
   </par>
   <par dur="30min">
    <audio src="rtsp://content-server/content-A-3.mp4" />
    <video src="rtsp://content-server/content-V-3.mp4" region="v" />
   </par>
  </seq>
 </body>
</smil>
```

Fig. 5 Modified SMIL file after content segmentation.

(EX-1006, Fig. 5; *id.*, 1780-81; EX-1002 ¶44.) This modified SMIL file identifies files for different content streams from the same work, for example audio (content-**A**.mp4) and video (content-**V**.mp4) files. (EX-1006, 1780-81 (defining "content names"), 1782.) The "modified SMIL files indicate the names, durations, and relations of the segments." (*Id.*, 1781.) For example, each segment is played sequentially for the duration defined by the "<par dur>" entry. (*Id.*) Thus, the modified SMIL file is a "descriptor file" at least because it identifies and provides timing information for content. (EX-1002 ¶44; EX-1001, 3:6-11 (descriptor file "defines a plurality of digital media files" using time information), Abstract, 18:27-41, 19:26-20:9, Figs. 5c, 6, 20, 22, Table 1.)

b. Media Streams

Yoshimura's portal servers also store, for example, two streams: (a) segments of "'content-A.mp4' for audio content"; and (b) segments of "'content-V.mp4' for video content." (EX-1006, 1780-81; EX-1002 ¶45.) Yoshimura's server divides the original audio file into smaller files and stores the segments on the portal server before sending the segments to a content server for storage. (EX-1006, 1780; *id.*, 1780 (content stored on surrogate servers), 1782-83; EX-1002 ¶45.)

c. Same Originating Work

i. Yoshimura

Yoshimura distributes "streaming content" including audio and video streams. (EX-1006, 1779-81; EX-1002 ¶46.) It would have been obvious that those streams would be derived from the same written work. (EX-1002 ¶46.) For example, it would have been obvious to use Yoshimura to distribute audio and video streams of movies or television, which would be derived from the same written work—a script. (*Id.*; *see also* EX-1001, 18:44-45; EX-1098, 5 (PO asserting audio and video of a movie originates from the same written work (script)).) It would have also been obvious that Yoshimura's media streams would include audio with images or text from a book, as audiobooks were widely known. (EX-1002 ¶46.) Such streams were derived from the same written work—a book. (*Id.*) Indeed, a POSITA would have understood Yoshimura to disclose streaming all media content supported by SMIL, e.g., text, audio, and images. (EX-1003, vii, 3; EX-1012, 108, 123; EX-1002 ¶46.)

ii. Lindahl

Lindahl discloses a system for transmitting audiobooks using a network. (EX-1004 ¶¶[0001], [0010], [0015]-[0017], [0019]-[0024], [0030], [0043]-[0045], claims 1-12, Fig. 1.) The audiobooks include multiple media streams (e.g., audio/text) derived from the same written work. (*Id.* ¶¶[0004], [0001]; EX-1002 ¶47.)

Thus, Yoshimura, alone or with Lindahl, renders obvious at least one server (e.g., content provider, portal, and/or cache/surrogate servers) having stored thereon a descriptor file (e.g., modified SMIL file) and a plurality of media streams (e.g., audio/video/text) derived from a same originating written work (e.g., script/book). (EX-1002 ¶¶43-48.)

iii. Motivation to Combine

A POSITA would have been motivated to transmit audiobooks (as in Lindahl) using Yoshimura's system. (*Id.* ¶¶49-53.)

First, Yoshimura teaches transmitting media content using SMIL files to "enhance[] streaming media quality … while utilizing network recourses efficiently." (EX-1006, 1778.) It improves caching by dividing files into smaller segments and allows for "easily manag[ing] the timing and spatial relations among these segments." (*Id.*, 1779.) It also "enables efficient pre-fetch," can be applied to any XML-related technology, is 3G compliant, and is easily implemented. (*Id.*) A POSITA would have been motivated to use Yoshimura's approach for transmitting audiobooks to obtain these benefits. (EX-1002 ¶50.)

Second, SMIL files, as in Yoshimura, were routinely used to transmit audiobooks. (EX-1004 ¶¶[0002]-[0003] (audiobooks transmitted using DAISY); EX-1003, vii, 3, 6-7, 21 (DAISY audiobook standard based on SMIL); EX-1046, 432; EX-1002 ¶51.)

Third, because audiobooks were widely known and supported, transmitting them using Yoshimura's system would have been a simple substitution of one known element/technique (Lindahl's audio/text streams) for another (Yoshimura's audio/video streams) in a known device ready for improvement to obtain predictable results (improved delivery of audiobooks). (EX-1004 ¶¶[0002]-[0003], [0005]; EX-1002 ¶52); *KSR*, 550 U.S. at 417.

A POSITA would have reasonably expected success in transmitting audiobooks using Yoshimura's system at least because Yoshimura already uses "modified SMIL files" to "enable multimedia content to be delivered" (EX-1006, 1778), and because using SMIL files for audiobooks was widely known (EX-1004 ¶[0004]; EX-1003, vii, 3). (EX-1002 ¶53.)

4. Element-1[c]

Element 1[c] recites "each media stream including a plurality of digital data files."

Yoshimura discloses "segmenting" the media content on its portal server into multiple "smaller segments." (EX-1006, 1780, 1782-83, Fig. 3; EX-1002 ¶56.) Each segment is stored as a separate file, e.g., files "content-A-1.mp4," "content-A-2.mp4," etc. for audio and files "content-V-1.mp4," "content-V-2.mp4," etc. for video. (EX-1006, 1781, Fig. 5; EX-1002 ¶56.) Accordingly, Yoshimura discloses or renders obvious each media stream (e.g., audio, video, text) including a plurality

of digital data files (e.g., content-A-{1,2,3}.mp4 and content-V-{1,2,3}.mp4). (EX-1002 ¶¶55-57.)

5. Element-1[d]

Element 1[d] recites "the descriptor file including information allowing a simultaneous synchronized rendering of the plurality of media streams that provides a literary experience of the originating written work."

Yoshimura's SMIL file includes information describing the synchronization of media streams and specifying how and when the content is displayed (e.g. Yoshimura's <par> and <seq> elements). (EX-1006, 1780-81, Fig. 5, 1779 (SMIL file describes "media synchronization about the content."), 1780 (SMIL file describes "timing relations among these segments"); EX-1051, 123; EX-1012, 137-38; EX-1002 ¶59.) A POSITA would have understood that Yoshimura's audio/video (e.g., movie) content and Lindahl's audio/text (e.g., audiobook) content provide a literary experience of a script or book. (EX-1002 ¶59.)

Thus, Yoshimura, alone or with Lindahl, discloses or renders obvious that the descriptor file (modified SMIL file) includes information allowing a simultaneous synchronized rendering of the plurality of media streams (e.g., audio narration/video/text) that provides a literary experience of the originating written work (e.g., script/book). (*Id.* ¶¶58-60.)

6. Element-1[e]

Element 1[e] recites "the information including time information for each digital data file in each media stream, the time information for synchronizing the plurality of media streams and determined relative to a timeline of an audio recording of the originating written work."

Yoshimura's content files are grouped within the SMIL file into a series of parallels ("par" or "<par>"). (EX-1006, 1781.) This content is rendered in parallel. (EX-1012, 137-38; EX-1002 ¶62.) For example, audio "content-A-1.mp4" and video "content-V-1.mp4" will be rendered at the same time:

```
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
"http://www.w3.org/2001/SMIL20/SMIL20.dtd">
<smil xmlns="http://www.w3.org/2001/SMIL20/Language">
 <head>
  <layout>
   <region id="v" top="5" left="5" width="180" height="180"/>
  </layout>
 </head>
 <body>
  <seq>
   <par dur="10min">
    <audio src="rtsp://content-server/content-A-1.mp4" />
    <video src="rtsp://content-server/content-V-1.mp4" region="v" />
   </par>
   <par dur="20min">
    <audio src="rtsp://content-server/content-A-2.mp4" />
    <video src="rtsp://content-server/content-V-2.mp4" region="v" />
   </par>
   <par dur="30min">
    <audio src="rtsp://content-server/content-A-3.mp4" />
    <video src="rtsp://content-server/content-V-3.mp4" region="v" />
   </par>
  </seq>
 </body>
</smil>
```

(EX-1006, 1781; EX-1002 ¶62.)

The parallels are located inside a sequence ("seq" or "<seq>") and are played in the order they appear in the sequence. (*Id.*; EX-1012, 135, 154.)

The SMIL file also includes the duration ("dur") of each parallel. (EX-1006, 1781; EX-1002 ¶64.) In this example, the first parallel is 10 minutes long, the second 20 minutes, and the third 30 minutes:

```
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
"http://www.w3.org/2001/SMIL20/SMIL20.dtd">
<smil xmlns="http://www.w3.org/2001/SMIL20/Language">
 <head>
  <layout>
   <region id="v" top="5" left="5" width="180" height="180"/>
  </layout>
 </head>
 <body>
  <seq>
1 <par dur="10min">
    <audio src="rtsp://content-server/content-A-1.mp4" />
    <video src="rtsp://content-server/content-V-1.mp4" region="v" />
   </par>
2 <par dur="20min">
    <audio src="rtsp://content-server/content-A-2.mp4" />
    <video src="rtsp://content-server/content-V-2.mp4" region="v" />
   </par>
3 <par dur="30min">
    <audio src="rtsp://content-server/content-A-3.mp4" />
    <video src="rtsp://content-server/content-V-3.mp4" region="v" />
   </par>
  </seq>
 </body>
</smil>
```

(*Id.*) The entire stream is the sum of its parallels/segments (here, 60 minutes long). (*Id.*) This time information in Yoshimura's SMIL file is determined relative to a timeline of media content (e.g., 0-60 minutes). (EX-1002 ¶64; *see* EX-1012, 125, 127, 141-47.)

Thus, Yoshimura, alone or with Lindahl, renders obvious time information (<seq>, <par>, duration, etc.) for each digital data file (e.g., content-A-{1,2,3}.mp4 and content-V-{1,2,3}.mp4) in each media stream (e.g., audio/video), the time information for synchronizing the streams and determined relative to a timeline (e.g., 0-60 minutes) of an audio recording of the originating written work (script/book). (EX-1002 ¶65.)

7. Element-1[f]

Element 1[f] recites "determining a first digital data file from the plurality of digital data files in a first media stream from the plurality of media streams, the first digital data file having digital content to be rendered with the media player, and determined using the time information in the descriptor file and a predetermined time offset, the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording." Predetermined time offsets include bookmarks, fast-forward/rewind requests, search results, etc. (EX-1001, claim 18.)

a. Yoshimura

Yoshimura determines media files in the same manner disclosed in the '720 patent (e.g., using a descriptor to process a user's fast-forward or rewind request). (EX-1002 ¶67.)

Yoshimura's media files can be determined based on their position within the body of a SMIL file (time information in the descriptor file). (EX-1002 ¶68; EX-1012, 85, 125, 135.) For example, at any given time in the presentation, Yoshimura's client will determine, request, and render the media file indicated by the parallel active at that time. (EX-1006, 1780, 1782, 1785; EX-1002 ¶68.)

Like the '720 patent, Yoshimura discloses the user may skip ahead or go back to previously viewed segments. (EX-1006, 1783.) To skip ahead to 15 minutes (predetermined offset) into the presentation (e.g., using fast-forward button or slider), a POSITA would have understood that Yoshimura's client will evaluate the start/end points for each parallel to find the parallel containing the requested time (15 minutes). (EX-1002 ¶69 (citing EX-1012, 167).) The client then determines and retrieves the media files associated with that parallel before rendering the file(s) starting at the requested time. (*Id*.) Thus, Yoshimura discloses or renders this limitation obvious. (EX-1002 ¶¶67-69.)

b. Lindahl

This limitation also would have been obvious over Lindahl. Lindahl's audiobooks are "navigable" using a bookmark that "contains information on the time at which the user paused play." (EX-1004 ¶¶[0004], [0035], Fig. 2.) The time in the bookmark is "an offset relative to the beginning of the book." (*Id.* ¶[0035].) The

bookmarks are "cross-correlated with the corresponding portion" of the content "that is resident on the server." (*Id.* ¶[0036], Fig. 2.)

Lindahl's client obtains a bookmark from the server and the server determines "the starting point for play … as indicated by the selected bookmark." (*Id.* ¶[0040]; *id.* ¶¶[0031], [0033], Fig. 3.) The relevant audio file is then determined and sent to the user. (*Id.* ¶¶[0040], [0031], [0033], Fig. 3, claim 1.) This audio file can be one of many audio files for the audiobook. (*Id.* ¶¶[[0022] ("selective download of an audio file of a chapter of a selected book"), [0010], [0023], [0040], [0045], claim 1; EX-1002 ¶71.) Thus, Lindahl discloses a bookmark including a time offset used to determine which audio file to stream. (EX-1002 ¶71.) It would have been obvious that the clients of the Yoshimura-Lindahl system would determine the file by comparing Lindahl's offset to Yoshimura's SMIL file to identify the segment based on the sequence and duration of the segments. (*Id.*) For example, if the time offset was at 23 minutes, the system would identify the segment that starts at 10 minutes and ends at 30 minutes based on the duration and sequence information in the SMIL file. (*Id.*)

Yoshimura, alone or with Lindahl, therefore renders obvious determining a first digital data file (e.g., Content-A-2.mp4) from the plurality of digital data files (e.g., content-A-{1,2,3}.mp4) in a first media stream (e.g., audio stream) from the

plurality of media streams, the first digital data file having digital content (e.g., audio) to be rendered with the media player (e.g., client), and determined using the time information in the descriptor file (e.g., <seq>, <par>, and/or <dur>) and a predetermined time offset (e.g., skip to or bookmark value), the predetermined time offset external to the descriptor file (e.g., chosen by user) and determined relative to the timeline of the audio recording (e.g., 15 minutes into 60 minute presentation). (*Id.* ¶¶66-72.)

c. Motivation to Combine

A POSITA would have been motivated to determine a digital data file using a time offset from a bookmark as taught by Lindahl and time information in Yoshimura's SMIL file for several reasons. (*Id.* ¶¶73-76.)

First, Yoshimura teaches users can pause and later resume playback of media content described by the SMIL file. (EX-1006, 1783; EX-1002 ¶74.) Lindahl discloses how to use a bookmark with a time offset to accomplish this in an audiobook. (EX-1004 ¶¶[0035]-[0036], [0040]-[0041]; EX-1002 ¶74.) A POSITA would have recognized that using Lindahl's bookmark beneficially allows a user to access an audiobook at an arbitrary position, such as where the user previously stopped listening. (EX-1004 ¶[0035]; EX-1005 ¶[0005] (users wish to restart where they left off); EX-1002 ¶74.)

Second, using time information in SMIL files and bookmarks for playback of media content was well known. (*Supra* §§II.B-II.C; EX-1004 ¶¶[0035]-[0036]; EX-1002 ¶75.) Thus, using Lindahl's bookmark time offset and the Yoshimura's time information to determine the relevant audio segment represents a simple substitution of one known element/technique (bookmarked time offset) for another (skip ahead/back value) to obtain predictable results (determining a file using a bookmarked time offset and time information in a SMIL file). (*Id.*)

A POSITA would have reasonably expected success in combining these teachings at least because both Yoshimura and Lindahl are adapted for use with SMIL-file based systems. (EX-1006, 1778; EX-1004 ¶[0004]; EX-1003, vii, 3; EX-1002 ¶76.)

8. Element-1[g]

Element 1[g] recites "downloading the first digital data file or a digital data segment contained in the first digital data file and determined using the time information in the descriptor file from the at least one server so that the digital content is resident with the media player."

Yoshimura's client devices download the determined segment (i.e. media file or portion thereof) from servers. (*Supra* §VII.A.2VII.A.7; EX-1006, 1779 ("clients are receiving streaming segments"), 1781, 1785 ("downloading the requested content"), 1782 (server delivers "media packets that include audio and video payloads

to the mobile client"), 1784; EX-1002 ¶79.) Downloaded segments would be resident on the media player (in memory or storage). (EX-1002 ¶79.) Thus, Yoshimura discloses or renders this limitation obvious. (*Id.* ¶¶78-80.)

9. Element-1[h]

Element 1[h] recites "rendering the digital content using the media player at an arbitrary point determined using the predetermined time offset." Yoshimura's client devices play media content. (EX-1006, 1778 ("playback" of content), 1785 (playback on mobile clients); EX-1002 ¶82.) The playback begins at the point indicated by the time offset obtained via a resume/skip command or bookmark. (*Supra* §VII.A.7; EX-1002 ¶82.) Accordingly, Yoshimura, alone or with Lindahl, renders this limitation obvious. (EX-1002 ¶¶81-84.)

B. Claim 18

Claim 18 depends from claim 1 and recites "wherein the predetermined time offset is provided by one of a bookmark, a search result, a table of contents, an index, a list of tables, a list of figures, footnotes, quotations, a list of illustrations, a fast-forward request, and a rewind request."

This claim would have been obvious in view of Yoshimura alone or with Lindahl. (*Supra* §VII.A.7; *see also* EX-1098, 84 (PO asserting claim 18 is satisfied by "a bookmark" or "Skip Intro" feature); EX-1002 ¶¶85-86.)

C. Claim 25

Claim 25 recites a system mirroring the limitations of the method of claim 1. Accordingly, claim 25 would have been obvious. (*Supra* §VII.A; EX-1002 ¶¶87-89.)

D. Claim 26

Claim 26 recites a computer with computer readable code that, when executed, performs the same limitations of the method of claim 1 and therefore would have been obvious. (*Supra* §VII.A; EX-1002 ¶¶90-92.)

VIII. GROUND 1B: CLAIMS 2-10, 14-15, AND 22 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND HECKERMAN.

A. Claim 2

Dependent claim 2 recites "wherein the plurality of media streams comprises an audio stream including a plurality of digital audio files, the plurality of digital audio files created by segmenting the audio recording in dependence upon natural language gaps so that when the plurality of digital audio files are rendered sequentially using the media player an experience of a contiguous audio stream is reproduced for a user without reassembling sequential digital audio files, and wherein the first digital data file is a first digital audio file from the plurality of digital audio files in the audio stream." The '720 patent discloses segmenting audio at natural language

gaps by "locat[ing] areas of silence or low decibel levels." (EX-1001, 5:67-6:7.) Heckerman discloses the same thing. (EX-1002 ¶94.)

Heckerman's system segments "audio and text data into multiple audio and data files" based on the "recognized silence." (EX-1030, claim 22; *id.*, abstract, 13:14-27, 13:30-32; EX-1002 ¶95.) Heckerman therefore discloses segmenting an audio recording in dependence upon natural language gaps. (EX-1002 ¶95.)

A POSITA would have been motivated to segment the Yoshimura-Lindahl-Heckerman media streams as taught by Heckerman for several reasons. (*Id.* ¶¶96-100.)

First, Yoshimura teaches dividing media streams into multiple files and Heckerman describes the benefits of doing so based on natural language gaps. (*Supra* §VII.A.4; EX-1030, Abstract, 12:65-13:34, claim 22; EX-1002 ¶97.) Heckerman explains its methods are advantageous for synchronizing audio and texts in audiobooks, like those in the Yoshimura-Lindahl combination. (EX-1030, 3:5-53.)

Second, the combination represents the simple substitution of one known element/technique (dividing audio files based on natural language gaps) for another (segmenting media streams generally) to obtain predictable results (segmentation based on natural language gaps). (*Id.*, 12:65-13:34, claim 22; EX-1063, 113-15; EX-1080 ¶¶[0026]-[0027]; EX-1002 ¶98.)

Third, this combination merely applies a known technique (segmenting media streams based on natural language gaps) to a known device and method (Yoshimura's system and segmentation) that is ready for improvement and yields predictable results. (EX-1002 ¶99.)

A POSITA would have reasonably expected success in applying Heckerman's segmentation method in Yoshimura's system because Yoshimura discloses segmenting media content and because segmenting content based on gaps in natural language was well known. (*Supra* §VII.A.4; EX-1030, 12:65-13:34, claim 22; EX-1080 ¶¶[0026]-[0027]; EX-1063, 113-15; EX-1002 ¶100.)

B. Claim 3

Claim 3 depends from claim 2 and recites "wherein the time information includes at least one of a start time, an end time, and a duration of each digital audio file in the plurality of digital audio files, the start and end times of each digital audio file determined relative to the beginning of the audio recording."

Yoshimura's time information includes a duration. (*Supra* §VII.A.6.) The start and end time of each parallel are also inherently disclosed in Yoshimura's SMIL file because they are derived from the sequence and duration of each parallel. (EX-1012, 127, 141 ("In general, begin+dur=end"), 125 (SMIL elements "determine the relative start and end times of document objects"), 126, 128, 142-47; EX-1002 ¶103.) For example, the start and end times of each file in the example are shown:

```
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
"http://www.w3.org/2001/SMIL20/SMIL20.dtd">
<smil xmlns="http://www.w3.org/2001/SMIL20/Language">
  <head>
    <layout>
      <region id="v" top="5" left="5" width="180" height="180"/>
    </layout>
  </head>
  <body>
    <seq>
1   <par dur="10min">      Start time: 0 minutes | End time: 10 minutes
      <audio src="rtsp://content-server/content-A-1.mp4" />
      <video src="rtsp://content-server/content-V-1.mp4" region="v" />
    </par>
2   <par dur="20min">      Start time: 10 minutes | End time: 30 minutes
      <audio src="rtsp://content-server/content-A-2.mp4" />
      <video src="rtsp://content-server/content-V-2.mp4" region="v" />
    </par>
3   <par dur="30min">      Start time: 30 minutes | End time: 60 minutes
      <audio src="rtsp://content-server/content-A-3.mp4" />
      <video src="rtsp://content-server/content-V-3.mp4" region="v" />
    </par>
    </seq>
  </body>
</smil>
```

(EX-1006, 1781; EX-1002 ¶103.) Even if the start and end time were not inherent in Yoshimura's SMIL file, it would have been obvious to add them because doing so was known in the art. (*Id.*; EX-1012, 141-42 (<par> elements can have begin/end attributes).) Yoshimura and Lindahl therefore disclose or render this limitation obvious. (EX-1002 ¶¶102-05.)

C. Claim 4

Claim 4 depends from claim 3 and recites "wherein determining the first digital data file comprises comparing the predetermined time offset to the at least one of a start time, an end time, and a duration of each digital audio file in the plurality

of digital audio files to find a first digital audio file that includes digital audio data from a point in the audio stream corresponding to the predetermined time offset."

Yoshimura discloses or renders obvious this limitation. (*Supra* §VII.A.7; EX-1002 ¶¶106-07.)

D. Claim 5

Claim 5 depends from claim 4 and first recites "wherein downloading the first digital data file or the digital data segment contained in the first digital data file and determined using the time information in the descriptor file comprises downloading the first digital audio file from the at least one server so that the first digital audio file is resident with the media player." Yoshimura discloses or renders obvious this limitation. (*Supra* §VII.A.8.)

Claim 5 further recites "wherein rendering the digital content using the media player comprises rendering the first digital audio file using the media player while simultaneously downloading subsequent digital audio files from the at least one server." This limitation refers to "prefetching," which was widely known. (EX-1002 ¶109.) Indeed, Yoshimura discloses prefetching content to a server. (EX-1006, 1782-83). A POSITA would have been motivated to have Yoshimura's client also pre-fetch content in the same manner to ensure smooth playback. (EX-1002 ¶109 (citing EX-1051, 71-73 (prefetching known for SMIL files makes playback "smoother"); EX-1012, 394-97).)

E. Claim 6

1. Element-6[a]

Claim 6 depends from claim 4. Element 6[a] recites "wherein the plurality of media streams further comprises a text stream including a plurality of electronic text files created by segmenting an electronic text of the originating written work."

This limitation would have been obvious over Yoshimura (*supra* §VII.A.3.c) and further in view of Lindahl, which discloses such content files for audio books include text files. (EX-1004 ¶¶[0004], [0008]; EX-1002 ¶113.) Additionally, Heckerman discloses segmenting audio and text files corresponding to the "same work, e.g., literary work." (EX-1030, 3:5-10, claim 22; *id.*, Abstract, 13:14-27 (segmenting text files), 13:30-32; EX-1002 ¶113.)

A POSITA would have been motivated to use Heckerman's segmented text files in Yoshimura's system and would have reasonably expected success in doing so. (*Supra* §VIII.A; EX-1002 ¶114.)

2. Element-6[b]

Element 6[b] recites "wherein the time information further includes at least one of a start time, an end time, and a duration of each electronic text file in the plurality of electronic text files, the start and end times of each electronic text file determined relative to the beginning of the audio recording." This limitation was disclosed or obvious. (*Supra* §VII.A.6; EX-1002 ¶117.) Specifically, it would have

been obvious that Yoshimura's modified SMIL file, when applied to an audiobook comprising audio and text files (as taught by Lindahl and Heckerman), would include time information (e.g., the same start, end, and duration information as the audio). (EX-1002 ¶117.) Those start and end times (e.g., 10 minutes, 30 minutes) are determined relative to the beginning of the audio recording (e.g., 0 minutes).

F. Claim 7

Claim 7 recites essentially the same limitations as claims 1, 4, and 6. Claim 7 would have been obvious in view of Yoshimura, Lindahl, and Heckerman for the reasons discussed for those claims. (*Supra* §§VII.A.7, VIII.C, VIII.E; EX-1002 ¶¶120-21.)

G. Claim 8

Claim 8 depends from claim 7 and recites "rendering the digital content using the media player comprises rendering the first digital audio file using the media player while simultaneously rendering the first electronic text file using the media player starting from a point in the first electronic text file calculated to correspond to the predetermined time offset."

This limitation was obvious in view of Yoshimura, Lindahl, and Heckerman. (*Supra* §§VII.A.3 (text and audio); VII.A.9 (rendered), VII.A.5 (simultaneous), VIII.C, VIII.D; EX-1002 ¶¶122-23.)

H. Claim 9

Claim 9 depends from claim 1 and recites essentially the same additional limitations as claims 6 and 7. This claim would have been obvious over Yoshimura, Lindahl, and Heckerman. (*Supra* §§VIII.E -VIII.F; EX-1002 ¶¶124-25.)

I. Claim 10

Claim 10 depends from claim 9 and recites essentially the same limitations as claim 4, but for a text file rather than an audio file. This limitation would have been obvious in view of Yoshimura, Lindahl, and Heckerman. (*Supra* §VIII.C; EX-1002 ¶¶126-27.)

J. Claim 14

Claim 14 depends from claim 10 and recites "wherein rendering the digital content using the media player comprises determining a current time offset using at least one of a discrete and a timed event." A "discrete event" can occur "through some user interaction" such as "the user pushing a button, tapping a mouse button or touch sensitive screen, performing a drag-and-drop operation, and performing a cut & paste operation." (EX-1001, 22:33-41; *see also* EX-1098, 73.)

Yoshimura's skip forward/backward commands are "discrete event[s]." (*Supra* §VII.A.7; EX-1002 ¶129.) Lindahl also discloses creating an offset when a user requests a bookmark. (*Supra* §VII.A.7; EX-1004 ¶¶[0035] ("a bookmark may be created if requested"), [0036]; EX-1002 ¶129.)

Thus, this claim would have been obvious in view of Yoshimura, Lindahl, and Heckerman. (EX-1002 ¶¶128-31.)

K. Claim 15

Claim 15 depends from claim 14 and recites rendering the audio and text files from the point in each stream corresponding to the predetermined time offset. This would result from combining the references above to render the audio and text files from, e.g., a bookmark or skip command. (EX-1002 ¶132.) Thus, this limitation was obvious in view of Yoshimura, Lindahl, and Heckerman. (*Supra* §§VII.A.9; VIII.A, VIII.J; EX-1002 ¶¶132-33.)

L. Claim 22

Claim 22 depends from claim 14 and recites "creating a bookmark for a selected point in the digital content, the bookmark including the current time offset, the current time offset determined relative to the timeline of the audio recording." This limitation would have been obvious in view of Yoshimura, Lindahl, and Heckerman. (*Supra* §§VII.B, VII.A.7, VIII.J; EX-1002 ¶¶134-35.)

IX. GROUND 1C: CLAIMS 5 AND 11-13 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND BULTERMAN.

A. Claim 5

Claim 5 depends from claim 4 and adds downloading and rendering the first digital audio file, "while simultaneously downloading subsequent digital audio files

from the at least one server." This would have been obvious over Yoshimura, Lindahl, and Heckerman. (*Supra* §VIII.D.) It would have been obvious also in view of Bulterman, which discloses clients prefetching content files while displaying other content. (EX-1012, 394-97; *id.*, 396 (rendering logo while simultaneously downloading subsequent logo file).) A POSITA would have been motivated to implement this feature in the Yoshimura-Lindahl-Heckerman combination because Bulterman teaches that prefetching allows rendering "to execute more efficiently." (*Id.*, 394; EX-1002 ¶139.) A POSITA would have reasonably expected success in implementing pre-fetching because it was well known and because Bulterman expressly teaches how to do so for SMIL files like Yoshimura's. (*Id.*)

B. Claim 11

1. Element-11[a]

Claim 11 depends from claim 10. Element 11[a] recites "wherein downloading the first digital data file or the digital data segment contained in the first digital data file and determined using the time information in the descriptor file comprises downloading the first electronic text file from the at least one server." This limitation would have been obvious. (*Supra* §VII.A.8; EX-1002 ¶142.)

2. Element-11[b]

Element 11[b] recites "loading the first electronic text file into a formatted memory structure, and loading a plurality of hard time tags from the descriptor file

into the formatted memory structure." A formatted memory structure "allows the identifiable parts of the text to be independently addressed." (EX-1001, 27:2-5; EX-1002 ¶143.) "Hard time tags" are "specifically determined" time offsets established for "identifiable points in the eText segments." (EX-1001, 26:1-29; EX-1002 ¶143.)

Yoshimura's SMIL file has start, end, and duration for each content file. (*Supra* §VII.A.6; EX-1002 ¶144.) These times serve as "hard time tags." (*Id.*)

Alternatively, Bulterman discloses <par> elements, like those in Yoshimura, can have "begin" and "end" attributes designating a particular time when the content in a parallel begins or ends. (EX-1012, 141-42, 145-47; EX-1002 ¶145.) The content in a parallel can include text. (EX-1012, 5, 13, 86, 128.) Begin and end attributes can be defined as a "temporal offset" against a "syncbase," which is the time some other SMIL element begins or ends. (*Id.*, 142-43 (begin time three seconds after syncbase), 140, 145-47, 311.) Bulterman teaches that parallels can be subdivided into second level parallels which can also contain begin and end attributes. (*Id.*, 143, 148 (outer and inner <par> elements).) A POSITA would have understood the begin and end attributes to be hard time tags of the text contained within a particular parallel. (Ex. 1002 ¶145.)

Yoshimura's content files (e.g., text file) and SMIL file are loaded into formatted memory when they are saved on the client device. (EX-1006, 1778 (SMIL

file downloaded to client), 1779 (clients receive segments and SMIL file); 1782; EX-1002 ¶146.) To the extent PO argues this limitation requires storing the time tags inside their corresponding content file, doing so would have been obvious. (EX-1002 ¶146.) A POSITA would have understood doing so was merely changing the content in a data file. (*Id.*) Such an adjustment would have been obvious to a POSITA. (*Id.*)

A POSITA would have been motivated to include begin and end attributes in Yoshimura's SMIL file because such a combination would simply be applying a well-known SMIL feature for its intended purpose (e.g., specifying when an object begins/ends). (*Id.* ¶147.) Such a modification represents the simple addition of known elements/techniques (begin/end attribute) to another element/technique (Yoshimura's SMIL file) to obtain predictable results (SMIL files with parallels begin/end attributes). (*Id.*) A POSITA would have reasonably expected success in applying Bulterman's timing attributes in Yoshimura's system because instructions for doing so were provided in Bulterman. (EX-1012, 141-47, 249; EX-1002 ¶147.)

Accordingly, Yoshimura in combination with Lindahl and Bulterman render obvious loading the first electronic text file into a formatted memory structure (e.g., client memory), and loading a plurality of hard time tags (e.g., begin/end) from the descriptor file (SMIL) into the formatted memory structure. (EX-1002 ¶¶143-48.)

Additionally, in its infringement contentions, PO argues this limitation is met by simply loading files into "computer memory" for "playback." (EX-1098, 62.) Yoshimura, Lindahl and Bulterman disclose loading files into computer memory for playback. (*Supra* §§VII.A.8-VII.A.9, VIII.H; EX-1002 ¶149.)

C. Claim 12

Claim 12 depends from claim 11 and recites "wherein rendering the digital content using the media player comprises determining at least one of a start time and an end time of a plurality of paragraphs in the loaded first electronic text file using the plurality of hard time tags, the start and end times of each paragraph determined relative to one of the beginning of the audio recording and a beginning of the first electronic text file."

Assigning start and end times to a plurality of text fragments would have been obvious in view of Yoshimura, Lindahl, and Bulterman. (*Supra* §IX.B; EX-1002 ¶153.)

It also would have been obvious that the text within each parallel of the Yoshimura-Lindahl-Heckerman-Bulterman combination would advantageously be a paragraph. (EX-1002 ¶154.) Heckerman also discloses text and audio files should be synchronized at the "start of each paragraph." (EX-1030, 3:17-19; *id.*, 2:59-63.) Bulterman discloses SMIL players can display paragraphs of text. (EX-1012, 86 (SMIL examples for text paragraph "Greetings!"); EX-1002 ¶154; *see also*

EX-1003, 21 (audiobook SMIL files synchronized at "paragraph, sentence, or even word level").) The start and end times of these paragraphs would then be determined using the begin/end attributes of Yoshimura or Bulterman. (EX-1002 ¶154.)

Additionally, as explained, Bulterman discloses begin/end attributes that are time offsets against a syncbase. (*Supra* §IX.B.2.) This syncbase can be the "the beginning or the end of some other element in the document" including the beginning of a <seq> element. (EX-1012, 140, 146, 147). Yoshimura' s content files are described in a single <seq> starting at the beginning of the entire content. (*Supra* §VII.A.6.) Thus, a POSITA would have understood the begin/end attributes can be offsets relative to the beginning of the audio content of Yoshimura. (EX-1002 ¶155; EX-1012, 328 (audio element can be "master media object for timing control purposes"). 347.) Annotated Figure 7-2 of Bulterman demonstrates this concept:

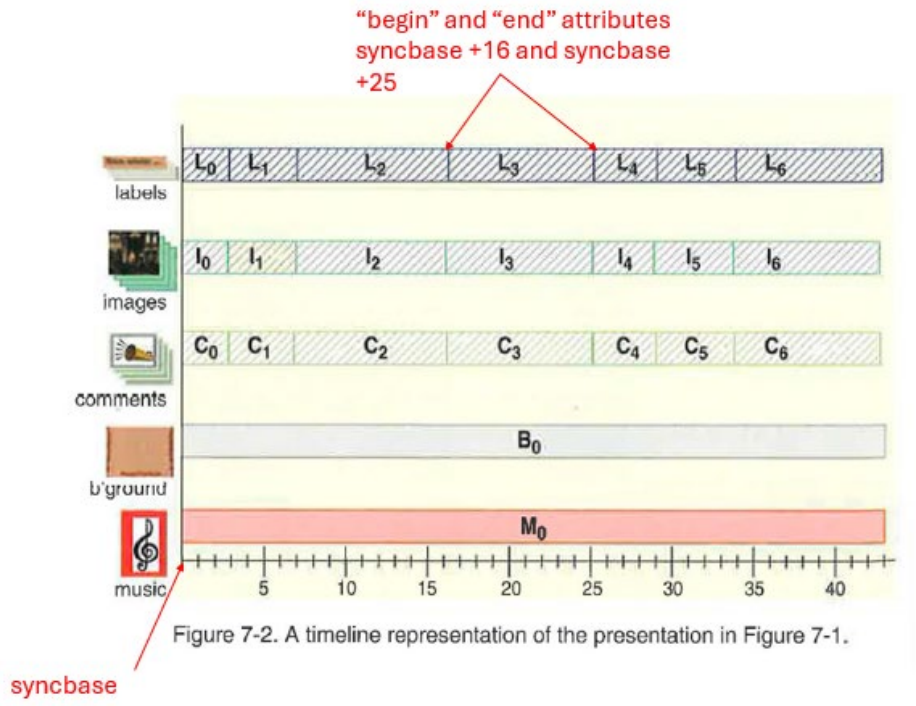


Figure 7-2. A timeline representation of the presentation in Figure 7-1.

(EX-1012, 128; EX-1002 ¶155.)

Therefore, Yoshimura, Lindahl, Heckerman, and Bulterman render obvious determining at least one of a start time and an end time of a plurality of paragraphs in the loaded first electronic text file using the plurality of hard time tags (e.g., begin/end), the start and end times of each paragraph determined relative to one of the beginning of the audio recording (e.g., syncbase set to beginning of recording) and a beginning of the first electronic text file. (EX-1002 ¶¶152-56.)

Additionally, in its infringement contentions, PO asserts this limitation is met by a "Media Presentation timeline and/or segment timing information." (EX-1098, 67.) Yoshimura, Lindahl and Bulterman disclose a media presentation timeline and/or segment timing information. (*Supra* §§VII.A.5-VII.A.6; EX-1002 ¶157.)

D. Claim 13

Claim 13 depends from claim 11 and recites "wherein rendering the digital content using the media player comprises extrapolating time values for a plurality of characters in the loaded first electronic text file, the time values extrapolated from the at least one of a start time, an end time, and a duration of the first electronic text file and the plurality of hard time tags."

Yoshimura and Bulterman teach applying hard time tags to text paragraphs in <par> elements, and Yoshimura teaches its <par> elements can have durations ("dur" values). (*Supra* §§IX.B.2, VII.A.6.) It would have been obvious to a POSITA that the combined system would extrapolate time values to the words and/or characters contained within the paragraph (e.g. plurality of characters) based on this information. (EX-1002 ¶161.) A POSITA would have been motivated to do so to provide time values for words or characters a user may wish to navigate to, such as the first word of a sentence or the punctuation character ending a sentence. (EX-1002 ¶161.) This would allow a user to navigate to particular times by searching for particular words or characters. (EX-1002 ¶161; *see also infra* §XIV.)

X. GROUND 1D: CLAIM 17 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND DTB.

Claim 17 depends from claim 1 and recites "wherein the plurality of media streams comprises an ancillary content stream including a plurality of ancillary files,

and wherein the time information includes at least one of a start time, an end time, and a duration of each ancillary file in the plurality of ancillary files, the start and end times of each ancillary file determined relative to the beginning of the audio recording, the ancillary content stream including at least one of a plurality of graphic files, a plurality of advertising files, and a plurality of user generated ancillary files."

DTB was the standard for distributing digital talking books like those described in Lindahl. (EX-1003; EX-1002 ¶165.) DTB's SMIL files synchronize audiobooks with graphic files, notes, annotations, and advertisements (e.g. ancillary content). (EX-1003, 26-27, 29-30, 163.)

A POSITA would have been motivated to include images and annotations in the Yoshimura-Lindahl audiobook content to diversify the types of content (e.g., children's books, comics, etc.) displayed by the system or allow a user to add annotations to take notes and otherwise mark-up the media content for future reference or to share with others. (EX-1002 ¶167; *see also* EX-1030, 3:5-55.) Such a combination represents the simple substitution of one known element/technique (synchronized media including images and/or annotations) for another (Yoshimura's streaming media) to obtain predictable results (streaming images and/or annotations). (*Id.*, 12:65-13:34, claim 22; EX-1063, 113-15; EX-1080 ¶¶[0026]-[0027]; EX-1002 ¶167.) A POSITA would have reasonably expected success in applying Hecker-

man's segmentation method in Yoshimura's system because streaming media including images and/or user annotations was well known. (EX-1002 ¶167; *supra* §II.B; *infra* §XVI.A.6.)

XI. GROUND 1E: CLAIM 20 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND DTB.

Claim 20 depends from claim 10 and recites "wherein the plurality of media streams further comprises an audio stream including a plurality of digital audio files created by segmenting the audio recording and an illustration stream including a plurality of illustration files from the originating written work, and wherein the descriptor file includes at least one of a start time, an end time, and a duration of each digital audio file in the plurality of digital audio files, and at least one of a start time, an end time, and a duration of each illustration file in the plurality of illustration files, the start and end times of each digital audio file and each illustration file determined relative to the beginning of the audio recording."

Yoshimura and Lindahl render obvious media streams including audiobooks. (*Supra* §VII.A.3; EX-1002 ¶170.) It would have been obvious an audiobook stream would include illustrations from the book segmented in the same manner. (*See, e.g.*, EX-1003, 60-61 (disclosing audiobooks with text and images are synchronized); EX-1002 ¶170.)

Additionally, DTB discloses SMIL-file-based media streams can include parallels comprising illustrations (e.g., images). (EX-1003, 26-27, 32-33; EX-1002 ¶171.) These parallels would have start, end, and durations in the same way disclosed by Yoshimura. (*Supra* §§VII.A.5-VII.A.8; EX-1002 ¶171.) Thus, claim 20 would have been obvious in view of Yoshimura, Lindahl, Heckerman and DTB. (EX-1002 ¶¶169-73.)

XII. GROUND 1F: CLAIMS 16 AND 24 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND HAY.

A. Claim 16

Claim 16 depends from claim 15 and recites "wherein rendering the digital content using the media player further comprises providing automated page turns within the first electronic text file using the current time offset."

Hay discloses "during audible reading of the narration by the software, when the reader software reaches the bottom of a page, the display will automatically advance to and display the next page" by means of a synchronization file linking the audio with the text. (EX-1028 ¶¶[0100], [0217]; EX-1002 ¶176.) Hay therefore discloses automatically turning from the first page to the second page in dependence upon the time information for the first page or for the second page (e.g., time associated with the bottom of the first page) as the portion of the audio stream is rendered (e.g., audio narration). (*Id.*; EX-1002 ¶176.)

A POSITA would have been motivated to automatically turn pages of Yoshimura's multimedia content (e.g., as disclosed by Hay) because this would improve the reader's experience, add convenience, and avoid unnecessary interruptions in the reading experience by eliminating the need to provide input at the end of the page . (EX-1002 ¶178.) Additionally, the combination represents the simple addition of one known element/technique (e.g., automatic page turning) to another known element/technique (e.g., multimedia presentations) to obtain predictable results (e.g., continuous rendering of content). (*Id.*); *KSR*, 550 U.S. at 417. Methods to automate page turning were commonplace and well-understood, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶178 (citing EX-1059, 49:6-7; EX-1057, Abstract; EX-1058, 4:32-40).)

B. Claim 24

Claim 24 combines the limitations of previously discussed claims. Claim 24 would have been obvious. (*Supra* §§VII.A, VIII.C-VIII.E, VIII.H-VIII.I, VIII.K, XII.A; EX-1002 ¶¶180-82.)

XIII. GROUND 1G: CLAIMS 19, 23, AND 27 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, AND SULL-948.

A. Claim 19

Claim 19 depends from claim 1 and recites "wherein the predetermined time offset is provided by a bookmark, the bookmark including information to identify

the descriptor file." Lindahl discloses providing a time offset from a bookmark. (*Supra* §VII.A.7; EX-1002 ¶185.)

Lindahl's bookmarks are linked to the specific book they identify. (EX-1004 ¶¶[0017], [0036], [0040].)

This limitation would have been obvious in view of Sull-948. Sull-948 discloses a bookmark file containing the bookmarked position within multimedia content and a "metadata identification (ID) of the bookmarked file." (EX-1060 ¶¶[0188]-[0192]; EX-1002 ¶186.) Sull-948's metadata includes description of "segments" of the multimedia content including temporal components of the audio-visual content. (EX-1060 ¶[0168]; EX-1002 ¶186.) Accordingly, Sull-948 discloses or renders obvious providing a predetermined offset (e.g., bookmarked position) by a bookmark, the bookmark including information to identify (e.g., metadata ID) the descriptor file (e.g., metadata describing segments). (EX-1002 ¶186.)

A POSITA would have been motivated to have Lindahl's bookmark identify a descriptor file as disclosed in Sull-948 because it would efficiently link the bookmark to the SMIL file so the client could retrieve the bookmarked file. (*Id.* ¶187.) A POSITA would have understood this would be a faster and more efficient way to retrieve the bookmarked file. (*Id.*) A POSITA would have reasonably expected success because bookmarks and descriptor files were well known and the inclusion

of additional information in the bookmark (e.g., information to identify the descriptor file) would have been routine. (*Id.*)

B. Claim 23

Claim 23 depends from claim 1 and recites "determining a current time offset while rendering the digital content" and "creating a bookmark for a selected point in the rendered digital content, the bookmark an electronic file including the current time offset and an identification of the descriptor file, the current time offset being a time difference from a beginning of the audio recording to a selected point in the audio recording, the selected point in the audio recording corresponding to the selected point in the digital content."

Lindahl discloses determining current time offset while rendering the digital content (EX-1004 ¶[0035], Fig. 2), creating a bookmark for the selected point (*id.*, *see also id.* ¶[0036]), where the bookmark includes the current time offset (*id.*), the current time offset being from a beginning of the audio recording to a selected point in the audio recording (*id.*), the selected point in the audio corresponding to the selected point in the digital content (*id.* ¶¶[0004], [0008]). (*Supra* §VII.A.7.) It would have been obvious to identify a descriptor file in Lindahl or Sull-948's bookmark. (*Supra* §VII.A.7.b.) The selected point in the audio would correspond to the selected point in the digital content. (*Supra* §VII.A.5-VII.A.7; EX-1002 ¶191.)

C. Claim 27

Claim 27 depends from claim 26 and further recites "determine a current time offset while rendering the digital content" and "create a bookmark for a selected point in the rendered digital content, the bookmark an electronic file including the current time offset and an identification of the descriptor file, the current time offset being a time difference from a beginning of the audio recording to a selected point in the audio recording, the selected point in the audio recording corresponding to the selected point in the digital content." Lindahl's bookmark is created while the content is rendering, when the user pauses. (*Supra* §VII.A.7; EX-1002 ¶194.) Claim 27 would have been obvious. (*Supra* §XIII.B; EX-1002 ¶¶194-95.)

XIV. GROUND 1H: CLAIM 21 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, HECKERMAN, AND AMIR.

Claim 21 depends from claim 10 and recites "wherein the predetermined time offset is provided from a text search result." Amir discloses a text search of text associated with audio and video ("speech transcripts") that returns "the matching spoken keywords" and allows playback of content "at the time offset corresponding to the match." (EX-1040, 210; EX-1002 ¶197.)

A POSITA would have been motivated to obtain a time offset from a text search of media content because text search in multimedia content was well known. (*See, e.g.*, EX-1048, Abstract, 8:15-44, Fig. 12; EX-1003, 18; EX-1040, 210-11;

EX-1002 ¶198.) Thus, using Amir's text search in Yoshimura's system represents a simple addition of one known element/technique (text searching) to another (Yoshimura's media player playing content) to obtain predictable results (obtaining a time offset from a text search of media content). (EX-1002 ¶198.) Such a combination would also advantageously allow a user to identify locations in an audiobook through text search in addition to other methods.

A POSITA would have also reasonably expected success because text searching was well known, and because the SMIL 2.0 Standard, which governs SMIL features and capabilities, allows for text searching. (EX-1051; EX-1002 ¶199.)

XV. GROUND 1I: CLAIMS 28 AND 29 WOULD HAVE BEEN OBVIOUS IN VIEW OF YOSHIMURA, LINDAHL, SULL-948, AND DTB.

A. Claim 28

Claim 28 depends from claim 27 and recites that the computer readable storage medium, "when executed by a computer, further causes said computer to electronically transfer the bookmark to another computer."

Lindahl's bookmark "may [] be uploaded to the server[.]" (EX-1004 ¶[0036].) Lindahl therefore discloses computer code configured to cause said computer (e.g., client device) to transfer (e.g., upload) the bookmark to another computer (e.g., server). (*Id.*; EX-1002 ¶204.)

To the extent PO argues the other computer must be a client computer, this

limitation would have been obvious in view of Lindahl. (EX-1002 ¶205.) In addition to uploading a bookmark to the server, Lindahl discloses transferring the bookmark from the server back to a client device. (EX-1004 ¶[0031], Fig. 2 ("Get Bookmark"); EX-1002 ¶205.) Lindahl also discloses multiple clients can access the system. (EX-1004 ¶[0043].) Accordingly, it would have been obvious in view of Lindahl that a client device would transfer a bookmark through the server to another client device associated with the user. (*Id.*; EX-1002 ¶205.) A POSITA would have been motivated to implement Lindahl so that a user can start an audiobook on one device (e.g., audio book device 310) and later resume the audiobook on another device (e.g., handheld computer or cell phone 310). (EX-1002 ¶205.)

Even if Lindahl did not disclose or render obvious this limitation, it would have been obvious in view of DTB. (*Id.* ¶206.) DTB's players set bookmarks and "export" them to other devices. (EX-1003, 45.) A POSITA would have been motivated to incorporate this feature into the combined system to allow users to resume audiobook playback across devices or transfer bookmarks from one user to another. Such a combination represents the addition of one known element/technique (e.g., transferring bookmarks between players) to another known element/technique (e.g., Lindahl's transferring bookmarks to a server) to obtain a predictable result (e.g., players transferring bookmarks to other players). (EX-1002 ¶206.) A POSITA would have reasonably expected success because such a combination was well-

known and commonplace. (*Id.* (citing EX-1062, 2:63-3:6; EX-1069, 3:14-28; EX-1005 ¶[0063]).)

B. Claim 29

Claim 29 depends from claim 27 and recites the computer readable storage medium, "when executed by a computer, further causes said computer to transfer one of the bookmark and a shortcut to the bookmark to one of a desktop, a window other than the window used for rendering the digital content, and an application other than the application used for rendering the digital content."

Lindahl alone or with DTB discloses or renders obvious transferring a bookmark from one client to another. (*Supra* §XV.A; EX-1002 ¶210.) Lindahl's client devices include handheld computers or audiobook devices. (EX-1004 ¶[0016], Fig. 1; EX-1002 ¶210.) A POSITA would have understood an obvious way to transmit a bookmark file from one handheld computer to another would be to download the file to whatever location on the computer the user selected. (EX-1002 ¶210). Routine places would be computer desktops (including by a shortcut), a folder ("a window other than the window used for rendering the digital content"), or a file management system ("an application other than the application used for rendering the digital content."). (*Id.*) Thus, Yoshimura and Lindahl render this claim obvious. (*Id.* ¶¶209-12.)

XVI. GROUND 2A: CLAIMS 1-15, 17-18, 20-22, AND 25-26 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, AND OPTIONALLY STELOVSKY.

A. Claim 1[2]

1. Preamble

Reid discloses "a method and system" for synchronizing media streams, navigating through them, and "simultaneously displaying to the user some or all of the streams at the position corresponding to the position selected in the navigated stream." (EX-1076, 4:9-18; EX-1002 ¶215.) If the preamble is limiting, Reid discloses it. (EX-1002 ¶¶214-16.)

2. Element-1[a]

Reid discloses a client connected to a server via a network. (EX-1076, 6:8-9, Fig. 1.) The client is capable of playing video/audio streams. (*Id.*, 6:19-23, 12:7-14, 20:29-21:5, 21:25-27, Figs. 1, 5.) Reid therefore discloses "providing a media player having access to at least one server via a network." (EX-1002 ¶¶217-19.)

3. Element-1[b]

Reid synchronizes a plurality of media streams along the same timeline ("common time base") using an index for each media stream. (EX-1076, 6:19-9:20, Figs. 2-4; EX-1002 ¶221.) The indices provide corresponding positions between the media streams to synchronize the streams during playback. (EX-1076, claim 5, 4:1-

[2] Claim language is not repeated here.

3, 12:19-26, 13:10-20, 20:21-21:3.) Reid's indices are a descriptor file. (EX-1002 ¶221.) If PO argues the descriptor file must be a single file, it would have been obvious to store the indices in a single file. (*Id.* ¶222; EX-1044, Fig. 2.) In fact, Reid refers to them collectively as an "identification means." (EX-1076, claim 5.)

The indices are stored on the server with the media streams. (*Id.*, 20:8-19 (indices stored on server and downloaded to client), 17:13-18 (media streams are stored on the server).)

Reid discloses the synchronized streams may be for a slide presentation, and the media streams may include audio, video, slides, and a transcript. (*Id.*, claims 7-14.)

Even if Reid did not render obvious media streams derived from the same originating work, it would have been obvious in view of Heckerman. (EX-1002 ¶225.) Heckerman discloses a computer system "for presenting books or other literary works stored in electronic form." (EX-1030, 5:59-65.) Heckerman synchronizes and segments "text and audio files corresponding to, e.g., the same exemplary literary work." (*Id.*, 5:34-58.) Heckerman discloses electronic distribution of books and other literary works via the internet. (*Id.* 1:32-41.)

Thus, Reid, alone or with Heckerman, renders obvious a server storing a descriptor file (e.g., identification means) and a plurality of media streams (e.g., audio,

text) derived from a same originating written work (literary work). (EX-1002 ¶¶220-26.)

A POSITA would have been motivated to use Reid's system to transmit Heckerman's audiobooks. (*Id.* ¶227.)

First, Reid teaches presentations on CD-ROMs suffered from limited navigation capabilities. (EX-1076, 2:12-17, 3:7-16; *see also* EX-1001, 2:18-25 (patentee admitting audiobooks were hard to navigate).) Reid explained its system overcame these problems and provided easy-to-navigate, synchronized digital programming. (EX-1076, 3:20-4:25.) A POSITA would have been motivated to use Reid's system to gain these benefits when providing Heckerman's audiobooks. (EX-1002 ¶228.)

Second, because audiobooks were widely known and supported, transmitting them using Reid's system would have been a simple substitution of one known element/technique (Heckerman's segmented audio book audio/text streams) for another (Reid's organization of audio/text streams) in a known device ready for improvement to obtain predictable results (organization of audiobook media streams). (EX-1004 ¶¶[0002]-[0003], [0005]; EX-1002 ¶229.)

A POSITA would have reasonably expected success in transmitting audiobooks using Reid's system at least because synchronized rendering of media files was widely known and commonly used for audiobooks. (EX-1004 ¶[0004] (audiobook "created using the DAISY format"); EX-1002 ¶230.)

4. Element-1[c]

This limitation would have been obvious in view of Reid alone or with Heckerman. (EX-1002 ¶232.) Reid's hypertext object and slide media streams are stored as separate files. (EX-1076, 8:2-3.) Heckerman teaches dividing all media streams in a presentation (e.g., text and audio streams) into a plurality of smaller files. (EX-1030, 13:14-34; EX-1002 ¶233.)

A POSITA would have been motivated to divide Reid's media streams into a plurality of files for multiple reasons. (EX-1002 ¶¶234-37.)

First, a POSITA would have recognized the benefits of doing so including: (1) efficient data streaming and storage; and (2) easier seeking, faster playback, and simplified network handling. (EX-1006, 1778-80; EX-1002 ¶235.)

Second, the combination represents the simple addition of one known element/technique (e.g., Heckerman's file segmentation) to another (e.g., Reid's media streams) to obtain predictable results (e.g., media streams comprising multiple files). (EX-1002 ¶236.).

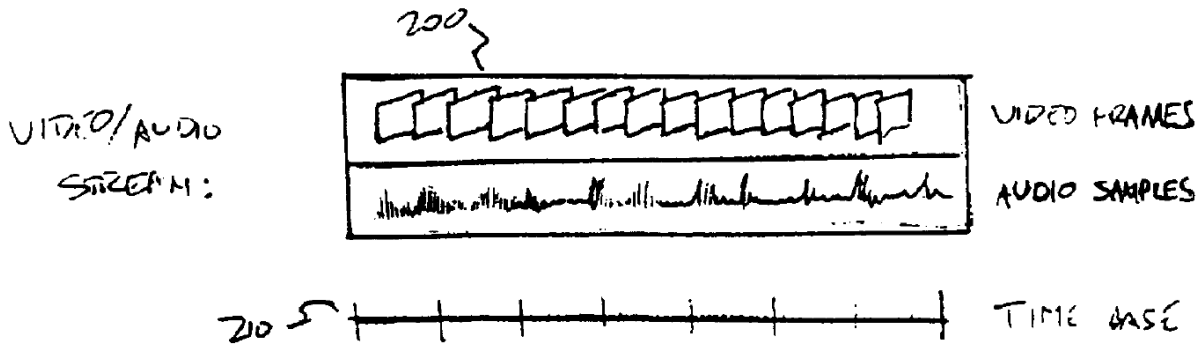
A POSITA would have reasonably expected success at least because storing media streams (including audio and text streams) as a plurality of files was commonplace. (EX-1002 ¶237 (citing EX-1004 ¶¶[0022]-[0023]; EX-1006, 1780-81).)

5. Element-1[d]

Reid's indices are used for simultaneous synchronized rendering of the media files. (EX-1076, 17:4-11 ("multiple synchronized streams of each data type" in "single multimedia program"), claim 1, 3:20-23, 4:9-18; EX-1002 ¶240.) Because Reid, alone or with Heckerman, renders obvious that the content would be an originating written work, the simultaneous rendering would provide a literary experience of that work. (*Supra* §XVI.A.3; EX-1002 ¶241.) Thus, this limitation was obvious in view of Reid and Heckerman (*Id.*)

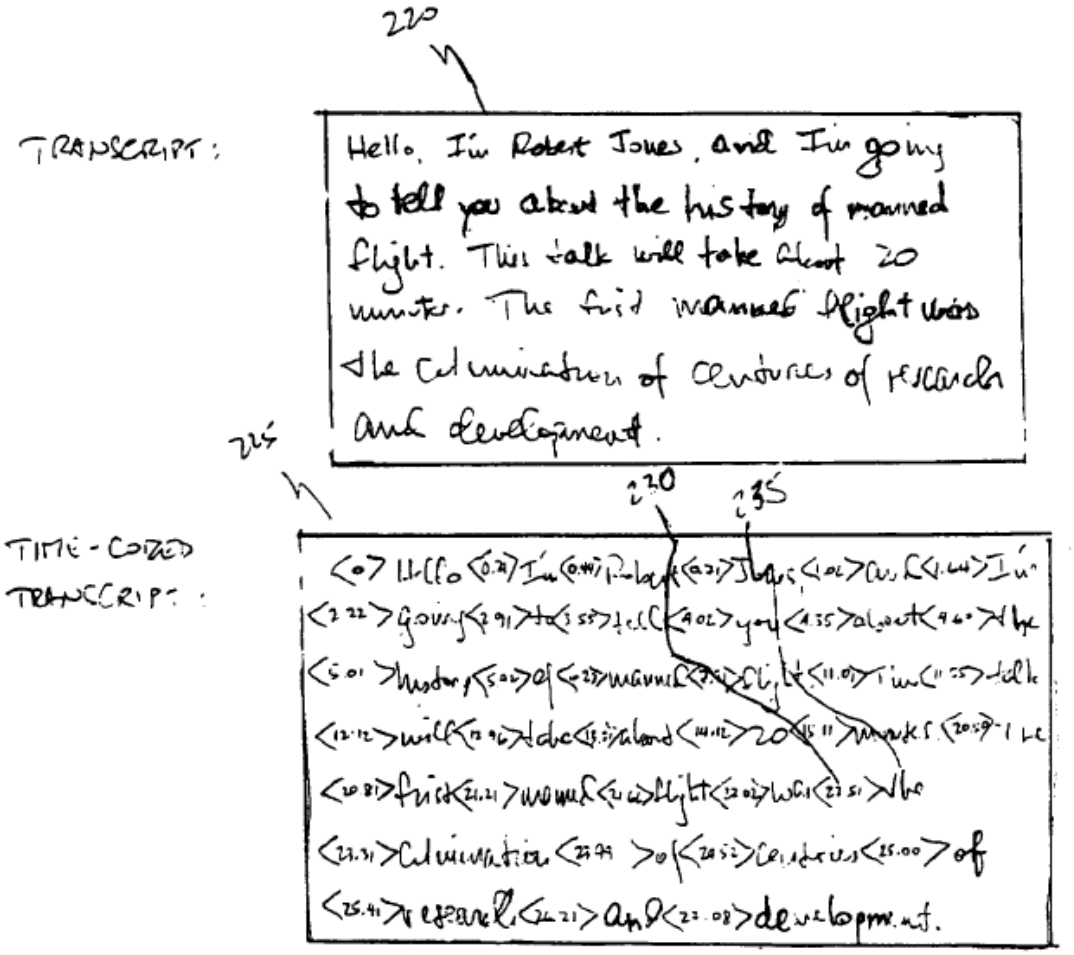
6. Element-1[e]

Reid's indices reference a "common time base" to synchronize media streams. (EX-1076, 4:9-18, Fig. 2, claims 1, 5.) Figure 2 shows the time base corresponds to a video/audio stream. (*Id.*, 16:28-30, 7:6-21, Figs. 2, 4; EX-1002 ¶243.)



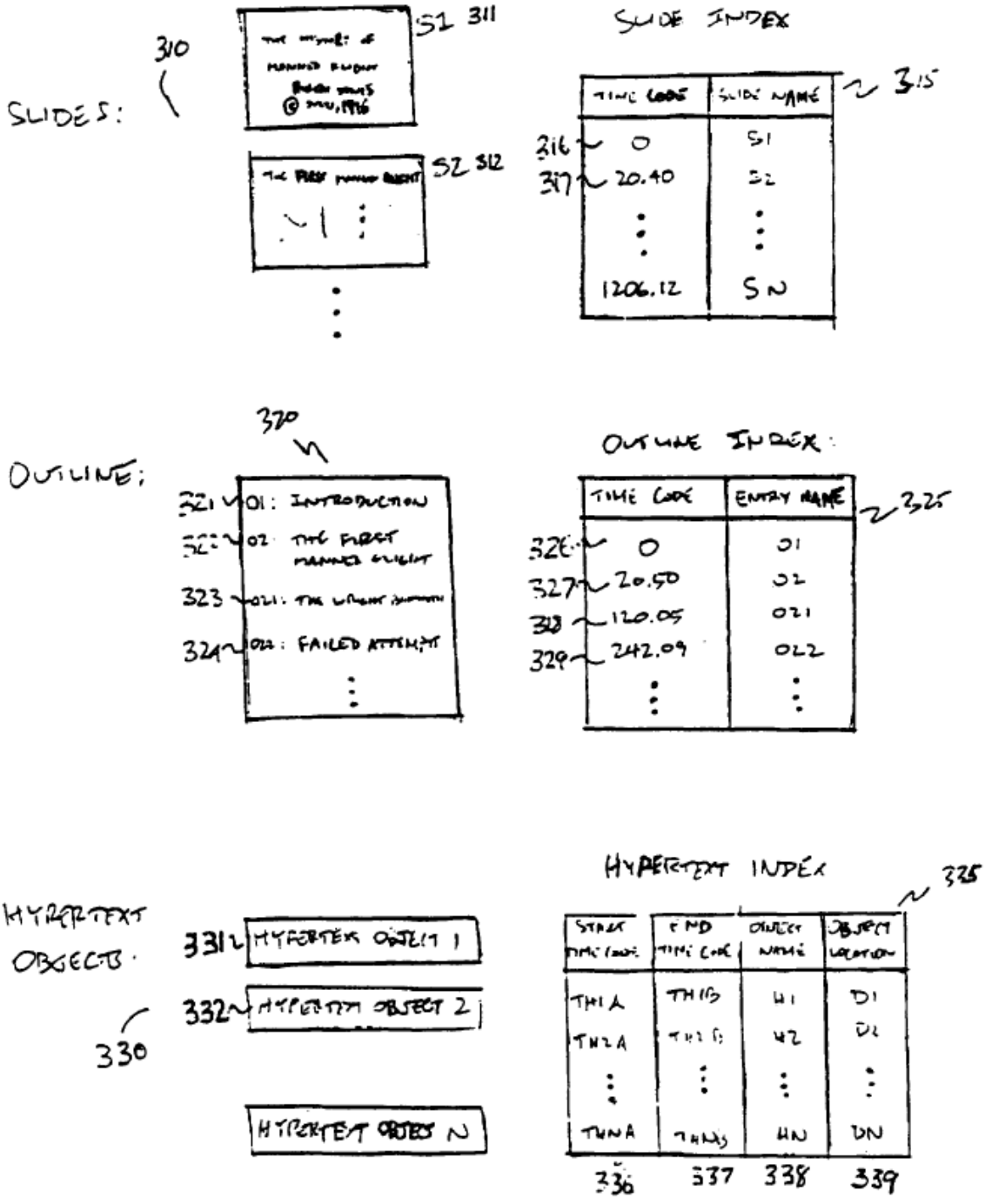
(EX-1076, Fig. 2 (excerpt); EX-1002 ¶243.)

A transcript is synchronized to the audio data stream by assigning a time code to each word, producing a time-coded transcript:

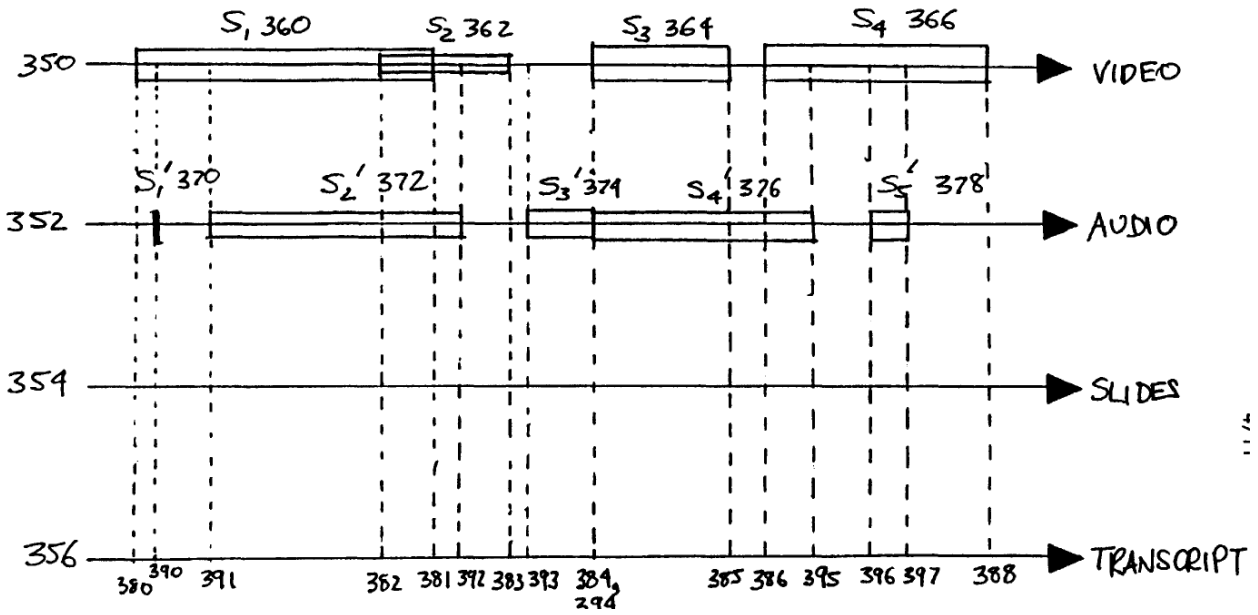


(EX-1076, Fig. 2 (excerpt).) Graphics (e.g., slides), outlines (table of contents), and hyperlinks are similarly assigned a timecode corresponding to the audio/video stream to create indices. (*Id.*, 7:23-9:14, Fig. 3.) These non-text indices are illustrated in Figure 3:

FIG. 3


SLIDES:
310
S1 311
S2 312
SLIDE INDEX
TIME CODE
SLIDE NAME
315
316
0
S1
317
20.40
S2
1206.12
SN
OUTLINE:
320
321
01: INTRODUCTION
322
02: THE FIRST MANNED FLIGHT
323
021:
324
022: FAILED ATTEMPT
OUTLINE INDEX:
TIME CODE
ENTRY NAME
325
326
0
01
327
20.50
02
328
120.05
021
329
242.09
022
HYPERTEXT OBJECTS:
330
331
HYPERTEXT OBJECT 1
332
HYPERTEXT OBJECT 2
HYPERTEXT OBJECT N
HYPERTEXT INDEX
335
START TIME CODE
END TIME CODE
OBJECT NAME
OBJECT LOCATION
TH1A
TH1B
H1
D1
TH2A
H2
D2
THNA
THNB
HN
DN
336
337
338
339

(*Id.*, Fig. 3.) Other data types (e.g., animations, images, etc.) can be synchronized in the same way. (*Id.*, 9:16-20.) The streams can then be organized into a single timeline used for synchronization and navigation. An exemplary compiled timeline is illustrated in Figure 4:



(*Id.*, Fig. 4; EX-1002 ¶244.)

Accordingly, Reid's identification means/indices have time information for each segment in each media stream (e.g., text, audio, video, graphics), the time information for synchronizing the plurality of media streams and determined relative to a timeline of an audio recording. (EX-1002 ¶245.) It would have been obvious to configure the media streams as a series of files. (*Supra* §XVI.A.4; EX-1002 ¶245.) In that case, time information for each segment of the stream would be time information for each file within that stream. (*Id.*) It also would have been obvious

that the audio recording was of the originating written work. (*Supra* §XVI.A.3; EX-1002 ¶245.)

To the extent PO argues the descriptor file must be a single file, it would have been obvious to include all of Reid's indices and timeline in a single file (e.g., Excel file). (EX-1002 ¶246; EX-1044, Fig. 2.) Reid and optionally Stelovsky therefore render this limitation obvious. (*Id.*)

7. Element-1[f]

Reid discloses the user may navigate to a point in the streams using a bookmark (EX-1076, 10:9-15), slide bars associated with the content streams (*id.*, 11:15-12:5), by jumping directly to a point of interest (e.g., using the outline) (*id.*, 12:16-17:2), or by navigating to content in the transcript (e.g., using a text search) (*id.*, 12:4-5, 17:18-18:7). (*Id.*, 22:3-5 (repositioning/navigating a program is "a common user behavior").) Once a position in the streams (e.g., audio stream) is selected, Reid's system computes the time code of that position and looks up the content associated with that time code using the indices. (*Id.*, 13:4-16:3.) For example, Reid can determine a slide file associated with that time code using the slide index. (*Id.*, 13:4-24, 8:2-4; EX-1002 ¶249.) Reid's client then displays the slide associated with that time. (*Id.*) Similarly, when a new time position is selected, Reid's client downloads a new segment/file of audio data from the server. (EX-1076, 24:4-12; EX-1002 ¶249.)

Accordingly, Reid discloses or renders obvious determining a first digital data file (e.g., slide or audio file) from the plurality of digital data files in a first media stream from the plurality of media streams, the first digital data file having digital content to be rendered with the media player, and determined using the time information in the descriptor file (e.g., slide index in identification means) and a predetermined time offset (e.g., time associated with user's selection), the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording (e.g., common time base). (EX-1002 ¶¶248-50.)

It would have been obvious that the same approach would be used to retrieve content from any media stream in the presentation. (*Supra* §XVI.A.4; EX-1076, 24:4-14; EX-1002 ¶251.) For example, for an audiobook, the same process would be used to retrieve a text file associated with the time offset. (*Supra* §§XVI.A.3-XVI.A.4; EX-1002 ¶251.)

8. Element-1[g]

Reid discloses that, after the user selects a program to play, the client downloads the non-isochronous data (e.g., slides, text, indices, etc.) from the server. (EX-1076, 20:8-27.) The non-isochronous data is then stored in disk cache on the client. (*Id.*) The audio file is downloaded in segments as the presentation is rendered. (*Id.*, 20:21-21:15; EX-1002 ¶254.)

To the extent PO argues the downloading needs to be in response to the determination step in element 1[f], that is also disclosed by Reid. For example, hypertext objects are determined and retrieved as the presentation is being rendered. (EX-1076, 20:16-19, 8:25-9:14, 20:16-19, 24:14-20; EX-1002 ¶255.) Similarly, Reid discloses downloading the audio segment in response to a new position being selected. (*Id.*, 24:4-12; *supra* §XVI.A.4; EX-1002 ¶255.) It would have been obvious to download the non-audio files in the same manner Reid downloads audio files (e.g., as the presentation progresses as opposed to a batch download at the beginning). (EX-1002 ¶255.)

Thus, downloading the first digital data file (e.g., audiobook text file) or a digital data segment contained in the first digital data file and determined using the time information in the descriptor file (e.g., index) from the at least one server so that the digital content is resident with the media player (e.g., client device) would have been obvious in view of Reid and Heckerman. (*Id.* ¶¶253-56.)

9. Element-1[h]

Reid's client renders the presentation from the user-selected portion of the presentation using the time base value of the new position. (EX-1076, 24:4-12, 22:3-5, 20:29-21:15, 12:16-16:28; EX-1002 ¶258.) Reid therefore discloses this limitation. (*Id.* ¶¶257-59.)

Thus, Reid, Heckerman, and optionally Stelovsky render claim 1 obvious. (*Id.* ¶¶214-59.)

B. Claim 2

Reid discloses an audio stream can be segmented for download. (EX-1076, 20:21-27; EX-1002 ¶261.) This limitation would have been obvious in view of Reid alone. (EX-1002 ¶261.) It also would have been obvious in view of Heckerman. (*Supra* §§XVI.A.4, VIII.A; EX-1002 ¶261.)

A POSITA would have been motivated to segment media streams based on natural language gaps as taught by Heckerman. (*Supra* §VIII.A; EX-1002 ¶262.)

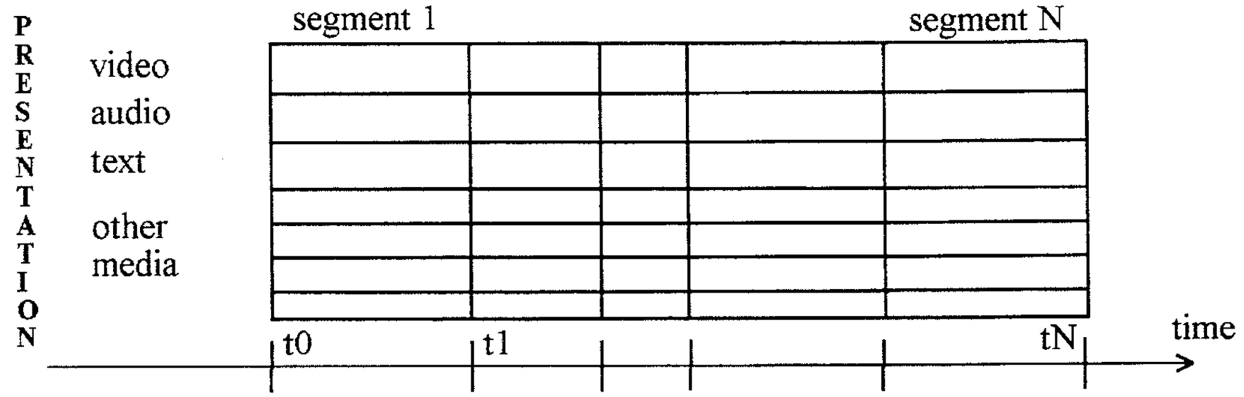
C. Claim 3

The Reid-Heckerman combination results in Reid's audio stream being divided into multiple audio files. (*Supra* §XVI.A.4; EX-1002 ¶266.) This limitation would have been obvious to a POSITA based on Reid's modified, segmented audio stream for at least two reasons. (*Id.*)

1. Audio Index

First, it would have been obvious to include an index for the audio files so that Reid's system would know which audio file to retrieve when a user navigates through the presentation. Such indices were known. (EX-1002 ¶267.) For example, in 1997, Stelovsky disclosed an index organizing a plurality of audio, video, text, and other media segments. (EX-1044, Abstract, 4:9-26, 3:26-67, 4:36-52, 5:20-25,

Fig. 2; EX-1002 ¶267.) The index, illustrated in Figure 2 below, is similar to the combined indices of Reid and includes an index for the audio file segments.



(*Compare* EX-1044, Fig. 2 (excerpt) *with* EX-1076, Fig. 4; EX-1002 ¶267.)

A POSITA would have been motivated to include an index of all media streams (including the audio stream) with start, end, and duration for each media segment/file for multiple reasons. (EX-1002 ¶¶268-71.)

First, a POSITA would have understood that having an index including audio segments/files would help Reid's system determine and navigate to the selected file. (EX-1044, 3:26-67 (index allows user to navigate multimedia stream); EX-1002 ¶269.) For example, when a user navigates within the presentation, Reid's "system computes the time base value of the new position" and downloads the segment of audio data associated with that position from the server. (EX-1076, 24:4-12, 10:9-16:25, 22:3-11; EX-1002 ¶269.) When the stream is divided into separate files (in-

stead of segments), a POSITA would have included an index for the audio files (instead of for the segments) to aid the system in determining the audio file (rather than segment) associated with the requested position. (*Id.*)

Second, the combination represents the simple addition of one known element/technique (e.g., Stelovksy's audio file index) to another (e.g., Reid's indices) in a known device ready for improvement to obtain predictable results (e.g., index identifying media files). (EX-1002 ¶270); *KSR*, 550 U.S. at 417.

A POSITA also would have reasonably expected success at least because indexing audio files in a stream was commonplace and because indexing files of an audio stream is no different than indexing other types of media/segments. (EX-1002 ¶271 (citing EX-1006; EX-1012; EX-1043; EX-1044).)

2. Shared Audio/Slide Index

Second, it would have been obvious that each of Reid's slides would correspond to a natural language break disclosed by Heckerman (e.g., a sentence or paragraph). (*Supra* §XVI.B; EX-1002 ¶272.) Thus, the start of each audio file would correspond to the start of each slide. (*Id.*; EX-1030, 3:17-25 (desirable to synchronize the start of each paragraph).) The slide index would serve as an index both for slides and the audio files. (*Id.*)

In both situations the start time provided for the audio file index would be relative to the beginning of the audio recording. (EX-1002 ¶273.) It also would

have been obvious to a POSITA in both cases to include an end time and duration in the index. (*Id*.)

Claim 3 is therefore obvious in view of Reid, Heckerman, and optionally Stelovsky. (*Id.* ¶¶265-75.)

D. Claim 4

Reid discloses this limitation in two ways.

First, as the user plays a presentation, Reid tracks the time of the audio and "[b]ased on the time code" of the audio stream, and displays the associated content. (EX-1076, 20:29-21:3, 4:11-25, 7:6-9:15; EX-1002 ¶277.)

Second, when the user "jumps" to a point in the presentation, it would have been obvious to look at the audio index, compare the time offset to the table values, and determine the file to download and play. (EX-1076, 12:16-15:28; EX-1002 ¶278.)

E. Claim 5

1. Element-5[a]

The Reid-Heckerman combination renders this limitation obvious. (*Supra* §§XVI.C-XVI.D, XVI.A.8.) A POSITA would have understood that, if a user jumped to a position in a different audio file, Reid would rely on its audio file index to determine the relevant audio file and the client would download it. (EX-1002 ¶281.)

2. Element-5[b]

When a user presses the play button, Reid downloads an audio segment (e.g., ~5-60 seconds) associated with the desired play location and stores the segment in memory cache on the client. (EX-1076, 20:21-21:15.) As the segments stream, "it depletes the memory cache." (*Id.*) When the amount of audio data in the cache falls below a threshold (e.g., 5-10 seconds), the client requests the next segment of data from the server. (*Id.*) Accordingly, Reid discloses or renders this limitation obvious. (*Id.*, *see also id.*, 23:20-26; EX-1002 ¶283.)

F. Claim 6

1. Element-6[a]

Reid and Heckerman render this limitation obvious. (*Supra* §§XVI.A.3-XVI.A.4, XVI.D; EX-1002 ¶285.)

2. Element-6[b]

A POSITA would have been motivated (and it would have been obvious) to divide Reid's transcript into files in view of Heckerman. (EX-1002 ¶287; *supra* §XVI.A.4.)

The Reid-Stelovsky index would reflect the start time of each file. (*Id.*) It also would have been obvious to include the end time and duration of the files in the index. (*Id.*)

G. Claim 11

1. Element-11[a]

Reid and Heckerman render this limitation obvious. (*Supra* §XVI.A.8; EX-1002 ¶290.)

2. Element-11[b]

Reid's "Time-Coded Transcript" has a hard time tag for each word. (EX-1076, 7:6-21, Fig. 2.) The time-coded transcript, and the other indices, are downloaded from the server to the client and stored in the client's memory. (*Id.*, 20:8-20.) The downloaded text file is loaded into a formatted memory structure at least because each word (e.g., time-coded) is identifiable and able to be independently addressed (e.g., rendered). (*Id.*, EX-1002 ¶292.) It would have been obvious to maintain time tags if the text stream were divided into larger text files than single-word files. (*Id.*) To the extent PO argues a formatted memory structure requires something more than the text contents being independently addressable, it would have been obvious to load the text file (and all other content files) into a formatted memory structure. (*Id.*; *supra* §IX.B.2.)

H. Claim 12

This limitation would have been obvious in view of Reid's hard time tags for multiple reasons. (*Supra* §XVI.G.) First, it would have been obvious that segmenting Reid's text steam into files according to the natural language gaps would include

segmenting into separate files for each paragraph. (EX-1030, 13:14-34, 3:18-25 ("When synchronizing text and audio versions of the same work, it is highly desirable to synchronize at least one point per paragraph, preferably at the start of each paragraph"); EX-1002 ¶296.) Thus, the indexed start and end time would serve as hard time tags for the paragraphs. (*Id.*) Second, to the extent the text files were larger than a paragraph, it would have been obvious to add hard time tags to the paragraphs for synchronization. (*Id.*; EX-1004 ¶[0022] (file can be "chapter of a selected book"); EX-1002 ¶297.) Third, regardless of whether the stream was divided into files by words, sentences, paragraphs, chapters, etc., a POSITA would have understood that, because each word in Reid has a time tag, the start and end of each paragraph would also necessarily have a time tag. (EX-1002 ¶298.) In each of these cases, these start/end times would be relative to the start of audio recording. (*Id.*)

I. Claim 13

To the extent PO argues this limitation requires extrapolating to a "plurality of characters" such as a word, it would have been obvious in view of Reid. For example, a POSITA would have been motivated to modify Reid's transcript index to only apply hard time tags to sentences, paragraphs, pages, or chapters instead of each word because it was known that word-level synchronization sometimes created problems. (EX-1091, 11-12 (word-level synchronization results in frequent losses

of synchronization between visual and audio modes); EX-1002 ¶302.) Thus, it would have been obvious for Reid's client to extrapolate time values for points in the text (e.g., words, sentences, etc.) that were in between the hard time tags. (*Id.*)

To the extent PO argues this limitation requires obtaining time values for the characters within a word, it would have been obvious in view of Reid. Reid synchronizes text content on a word-by-word basis. (EX-1076, 7:6-21, Fig. 2.) A POSITA would have been motivated to extrapolate the characters within a word to facilitate better, more exact highlighting or to improve text searching capabilities by jumping more directly to the location requested in Reid's text search. (*Id.*, 12:4-5, 17:13-18:7; EX-1002 ¶303.)

Thus, this limitation would have been obvious in view of Reid and Heckerman. (EX-1002 ¶¶301-05.)

J. Claims 14 and 18

Reid discloses multiple discrete events determining the time offset. (EX-1002 ¶307.) For example, Reid discloses navigating using bookmarks (EX-1076, 10:9-15); slide bars (e.g., outline slide bar, transcript slide bar, video slide bar, etc.) (*id.*, 11:15-12:5); selecting an outline entry (12:16-13:2); and text searching (*id.*; *see also id.*, 17:18-19:10). Reid therefore renders these limitations obvious. (EX-1002 ¶¶306-09.)

K. Claim 17

Reid discloses ancillary content streams including graphic files (e.g., slides, hypertext objects, or animations). (EX-1076, 9:15-20, 1:9-15, 2:24-28, Figs. 4-6; EX-1002 ¶311.) Each of these includes at least a start time. (*Id.*; *see also* EX-1076, Figs. 2-4.) It would have been obvious to also include an end time and duration for these objects. (*Id.*) Moreover, it would have been obvious to include graphic files, advertising, or user generated files in Reid's presentation. (*Id*.) For example, Stelovsky discloses using an index to synchronize user generated files with other media streams. (EX-1044, 3:61-67; EX-1002 ¶311.) Reid, alone or with Stelovsky, therefore discloses or renders obvious this limitation. (*Id.*)

L. Claim 22

Reid discloses creating bookmarks and using them to "return quickly to interesting parts of the presentation, or to pick up at the previous stopping point." (EX-1076, 10:9-15.) These bookmarks need not be created beforehand. (*Id.*, 10:7-14.) Thus, a POSITA would have understood that a user can create a bookmark, the bookmark including the current time offset (i.e., the current point in the book) relative to the common time base, which is the audio recording. (EX-1002 ¶315.)

Reid discloses that bookmarks are created using a similar approach to creating "conceptual events," and conceptual events may be "indexed" data streams like the

audio stream. (EX-1076, 10:9-11:14, Fig. 4.) Thus, a POSITA would have understood that Reid discloses that the bookmark's time offset is determined relative to the timeline of the audio recording. (EX-1002 ¶316.)

Reid therefore discloses or renders this limitation obvious. (*Id.* ¶¶314-18.)

M. Claims 7-10, 15, 20-21, and 25-26

Claims 7-10, 15, 20-21, and 25-26 recite similar limitations to the claims above and would have been obvious to a POSITA in view of Reid, Heckerman, and optionally Stelovsky. (*Supra* §§XVI.A.3-XVI.A.6 (text/audio/image streams/files and descriptor file for synchronized rendering), XVI.A.7-XVI.A.8 (text search, determining, downloading, and rendering), XVI.B-XVI.D (segmenting audio and start/end/duration), XVI.A.5 (simultaneous rendering); EX-1002 ¶¶319-36.)

XVII. GROUND 2B: CLAIMS 16 AND 24 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, HAY, AND OPTIONALLY STELOVSKY.

Hay discloses automatic page turning. (*Supra* §XII.A; EX-1002 ¶337.) A POSITA would have been motivated to present audiobook content (as disclosed in Heckerman) in a traditional page format to "benefit" readers. (EX-1003, vii; EX-1002 ¶338.) Thus, a POSITA would have been motivated to implement Hay's automatic page turning in Reid. (*Supra* §XII.A; EX-1002 ¶338.)

Claims 16 and 24 are therefore obvious in view of Reid, Heckerman, Hay, and optionally Stelovsky. (EX-1002 ¶¶337-40.)

XVIII. GROUND 2C: CLAIMS 19, 23, AND 27 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, SULL-948, AND OPTIONALLY STELOVSKY.

Bookmarks containing offsets (and being set based on the offset) would have been obvious to a POSITA. (*Supra* §§XVI.J, XVI.L; EX-1002 ¶345.)

Moreover, because Reid discloses viewers having access to multiple presentations at a time (EX-1076, 19:18-30), it would have been obvious to have bookmarks for multiple presentations. (EX-1002 ¶346.) In that case, a POSITA would have been motivated to include in Reid's bookmark an identifier of the associated index files so that the presentation (and the user's specific location) could quickly be identified and retrieved. (*Supra* §XIII.A; EX-1002 ¶346.)

Reid, Heckerman, Sull-948, and optionally Stelovsky therefore render claims 19, 23, and 27 obvious. (EX-1002 ¶¶341-47.)

XIX. GROUND 2D: CLAIMS 28 AND 29 WOULD HAVE BEEN OBVIOUS IN VIEW OF REID, HECKERMAN, SULL-948, DTB, AND OPTIONALLY STELOVSKY.

These claims would have been obvious in view of DTB. (*Supra* §XV; EX-1002 ¶348.) A POSITA would have been motivated to incorporate a transferrable bookmark with the claimed features. (*Supra* §XV; EX-1002 ¶349.) This is especially true given Reid's disclosure that a user can distribute programs to other users. (EX-1076, 19:24-30; EX-1002 ¶349.)

XX. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

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

XXI. DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE.

Efficiency, fairness, and the merits support institution. *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 (PTAB Mar. 20, 2020 ("*Fintiv*").

A. Factor 1: Potential Stay

On March 20, 2024, PO sued Petitioners for infringement of the '720 patent in the Eastern District of Virginia ("EDVA") in *Audio Pod IP, LLC v. Amazon.com, Inc.*, 2:24-cv-00185 (E.D. Va.) ("*Audio Pod I*"). On May 30, 2024 PO again sued Petitioners (except Audible, Inc.) for infringement of the '720 patent in *Audio Pod IP, LLC v. Amazon.com, Inc.*, 3:24-cv-00407 (E.D. Va.) ("*Audio Pod III*"), which was joined with a related case, *Audio Pod IP, LLC v. Amazon.com, Inc.*, 3:24-cv-00406 (E.D. Va.) ("*Audio Pod II*") (collectively, "the Litigations"). Petitioners will move to stay the Litigations pending resolution of this and related IPRs challenging the asserted patents. The EDVA routinely stays cases pending IPR proceedings, including pre-institution. *Sec. First Innovations, LLC v. Google LLC*, No. 2:23-cv-

00097, 2024 WL 234720 (E.D. Va. Jan. 22, 2024); *Sharpe Innovations, Inc. v. T-Mobile USA, Inc.*, No. 2:17-cv-00351, 2018 WL 11198604 (E.D. Va. Jan. 10, 2018).

On March 14, 2025, Petitioner Audible, Inc. ("Audible") filed a declaratory judgment action, seeking a declaration that Audible does not infringe the '720 patent, against PO in *Audible, Inc. v. Audio Pod IP, LLC*, 1:25-cv-02158 (S.D.N.Y.) (the "Audible DJ Action"). That action seeks a declaration of noninfringement only; the validity of the '720 patent is currently not at issue. However, Audible expects validity to become an issue in that case and, once it does, Audible expects to move to stay the Audible DJ Action in view of this IPR.

Thus, this factor weighs against denial.

B. Factor 2: Proximity of Trial to FWD

The EDVA has not set trial dates in the Litigations. The median time to trial in for civil cases in the EDVA for 2024 was 14.6 months[3], but it is clear a longer schedule will be needed here. (*See infra* §XXI.C.) The Audible DJ Action was filed recently, has no trial date, and SDNY's time-to-trial of 39.3 months is much longer than this proceeding. Thus, this factor weighs against denial. *Amazon.com, Inc. v.*

[3] *See* U.S. District Courts–Combined Civil and Criminal Federal Court Management Statistics (December 31, 2024), *available at* https://www.uscourts.gov/sites/default/files/2025-02/fcms_na_distprofile1231.2024.pdf.

Nokia Technologies OY, IPR2024-01140, Paper 9 at 9 (P.T.A.B. Feb. 12, 2025) (factor weighs against denial when there is no trial date); *Aptiv Services US, LLC v. Microchip Technology, Inc.*, IPR2024-00646, Paper 11 at 32 (P.T.A.B. Sept. 25, 2024) (same); *see Ericsson Inc. v. XR Communications LLC*, IPR2024-00613, Paper 9 at 34 n.12 (P.T.A.B. Oct. 9, 2024) ("median-time-to trial information" not useful where circumstances "do[] not reflect the normal course of a litigation").

C. Factor 3: Investment in Parallel Proceeding

PO filed its *Audio Pod I* complaint on March 20, 2024 and in *Audio Pod III* on May 30, 2024. Petitioners moved to dismiss or transfer *Audio Pod I* on May 31, 2024. No hearing date has been set for that motion. Otherwise, the parties have invested very little in *Audio Pod I*.

Fact discovery in *Audio Pod II* remains in its infancy, with only one deposition taken. The parties have not exchanged proposed claim constructions, submitted claim construction briefs or served expert reports. No case schedule has been set in *Audio Pod I*, and in *Audio Pod II*, the court's scheduling order does not extend beyond the *Markman* hearing on October 8, 2025. Thus, even if the Court denied Petitioners' pre-institution stay motion and the parties proceeded with claim construction over the coming months, much work will still remain, including expert reports, expert discovery, dispositive motions, pretrial motions, and trial.

Because the remaining investment in the Litigations and Audible DJ Action significantly outweighs any investment made thus far, this factor weighs against denial. *Samsung Electronics Co. v. Empire Technology Development LLC*, IPR2024-00896, Paper 15 at 13 (P.T.A.B. Dec. 13, 2024) (factor weighs against denial where *Markman* hearing and close of fact and expert discovery were after institution deadline); *Ericsson*, IPR2024-00613, Paper 9 at 34-35 (factor weighs against denial where "most efforts from the parties and court will take place after institution"); *Amazon.com*, IPR2024-01140, Paper 9 at 9-10.

D. Factor 4: Overlapping Issues

If this IPR is instituted and the Litigations and/or Audible DJ Action are stayed, Amazon could not pursue in the Litigations any invalidity ground raised or that could have been reasonably raised in this IPR. *Cal. Inst. of Tech. v. Broadcom Ltd.*, 25 F.4th 976 (Fed. Cir. 2022). If this IPR is instituted and the Litigations and/or Audible DJ Action are not stayed, Petitioners hereby stipulate not to pursue in those proceedings any ground of invalidity, against any claim challenged herein, that was raised or reasonably could have been raised in this Petition. Thus, this factor weighs heavily against discretionary denial. *Sotera Wireless, Inc. v. Masimo Corp.*, IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020).

E. Factor 5: The Parties

The parties are the same, but it is unlikely that the Litigations or Audible DJ Action will be tried before a FWD is entered in this IPR. Thus, this factor is neutral. *See Google LLC v. Jawbone Innovations, LLC*, IPR2022-00630, Paper 10 at 14 (P.T.A.B. Sept. 13, 2022).

F. Factor 6: Other Circumstances

The merits of this Petition are compelling, which favors institution. *Fintiv*, IPR2020-00019, Paper 11 at 18. Further, denying institution would negate Congress's intent in providing a 1-year period to file petitions and would encourage forum shopping as patent owners look to shield their patents from PTAB scrutiny by seeking judges with aggressive case schedules.

Thus, the Board should not decline institution under §314(a).

XXII. DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE.

The Office has not considered Yoshimura, Reid, DTB, Lindahl, Heckerman, Bulterman, Stelovsky, Amir, or Sull-948. Nor has the Office considered "substantially the same prior art or arguments." 35 U.S.C. §325(d). During prosecution, the Office considered Hay, but the Petition cites Hay for only two dependent claims alongside references the Office did not consider. This is sufficient to avoid denial. *Shenzen Chic Elecs. v. Pilot, Inc.*, IPR2023-00810, Paper 12 at 21 n.11 (P.T.A.B. Nov. 8, 2023) (denial inappropriate where challenges based on new art/arguments

"address all challenged claims"). The references here clearly disclose each claim element the Examiner thought was missing from the prior art. (*Supra* §III.B.) Accordingly, they are not—and could not be—cumulative of previously-considered references, absent material error by the Examiner. *See Quasar Sci. LLC v. Colt Int'l Clothing, Inc.*, IPR2023-00611, Paper 10 at 14 (P.T.A.B. Oct. 10, 2023). Thus, the Board should not deny institution under §325(d).

XXIII. CONCLUSION

Amazon requests institution and cancellation of all challenged claims.

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

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

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

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

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

PO asserted the '720 patent against Petitioners in patent infringement lawsuits captioned *Audio Pod IP, LLC v. Amazon.com, Inc. et al.*, No. 2:24-cv-00185 (E.D. Va., filed March 20, 2024) and *Audio Pod IP, LLC v. Amazon.com, Inc. et al.*, No.

3-24-cv-00407 (E.D. Va., filed May 30, 2024)[4]. Audible also filed a complaint for declaratory judgment of noninfringement of the '720 patent, captioned *Audible, Inc. v. Audio Pod IP, LLC*, No. 1:25-cv-02158 (S.D.N.Y., filed March 14, 2025). Petitioners further filed the following IPR petitions challenging claims of related patents:

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

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

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

Lead Counsel	Back-up Counsel
Colin B. Heideman (Reg. No. 61,513) 2cbh@knobbe.com BoxSEAZNL2185LP2@knobbe.com Postal and Hand-Delivery Address: KNOBBE, MARTENS, OLSON, & BEAR, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: (206) 405-2000 Facsimile: (206) 405-2001	Joseph R. Re (Reg. No. 31,291) 2jrr@knobbe.com Postal and Hand-Delivery Address: KNOBBE, MARTENS, OLSON, & BEAR, LLP 2040 Main Street, 14th Floor Irvine, CA 92614 Telephone: (949) 760-0404 Facsimile: (949) 760-9502 Christie R.W. Matthaei (Reg. No. 62,933)

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

Lead Counsel	Back-up Counsel
	2crw@knobbe.com Nathan D. Reeves (Reg. No. 77,806) 2ndr@knobbe.com Postal and Hand-Delivery Address: KNOBBE, MARTENS, OLSON, & BEAR, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: (206) 405-2000 Facsimile: (206) 405-2001 Daniel Hughes (Reg. No. 76,592) 2dph@knobbe.com Postal and Hand-Delivery Address: KNOBBE, MARTENS, OLSON, & BEAR, LLP 3579 Valley Centre Dr., Ste. 300 San Diego, CA 92130 Telephone: (858) 707-4000 Facsimile: (858) 707-4001

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

Please direct all correspondence to lead counsel and back-up counsel at the addresses shown above. Petitioners also consent to electronic service by email to BoxSEAZNL2185LP2@knobbe.com.

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

Petitioners certify the '720 patent is available for IPR and that Petitioners are not barred or estopped from requesting IPR on the identified grounds. This petition is being filed within one year of service of the original complaint against Petitioners in the district court litigation.

F. Payment of Fees (37 C.F.R. §42.15(a))

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

Respectfully submitted,

KNOBBE MARTENS OLSON & BEAR, LLP

Dated: March 24, 2025

By: */Colin B. Heideman /*

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

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

APPENDIX

Listing of Claims from U.S. 9,319,720	
Claim 1	
1[pre]	A method of rendering digital content comprising:
1[a]	Providing a media player having access to at least one server via a network,
1[b]	the at least one server having stored thereon a descriptor file and a plurality of media streams derived from a same originating written work,
1[c]	each media stream including a plurality of digital data files,
1[d]	the descriptor file including information allowing a simultaneous synchronized rendering of the plurality of media streams that provides a literary experience of the originating written work,
1[e]	the information including time information for each digital data file in each media stream, the time information for synchronizing the plurality of media streams and determined relative to a timeline of an audio recording of the originating written work;
1[f]	determining a first digital data file from the plurality of digital data files in a first media stream from the plurality of media streams, the first digital data file having digital content to be rendered with the media player, and determined using the time information in the descriptor file and a predetermined time offset, the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording;
1[g]	downloading the first digital data file or a digital data segment contained in the first digital data file and determined using the time information in the descriptor file from the at least one server so that the digital content is resident with the media player; and
1[h]	rendering the digital content using the media player at an arbitrary point determined using the predetermined time offset.

Listing of Claims from U.S. 9,319,720	
Claim 2	
--	A method of rendering digital content according to claim 1, wherein the plurality of media streams comprises an audio stream including a plurality of digital audio files, the plurality of digital audio files created by segmenting the audio recording in dependence upon natural language gaps so that when the plurality of digital audio files are rendered sequentially using the media player an experience of a contiguous audio stream is reproduced for a user without reassembling sequential digital audio files, and wherein the first digital data file is a first digital audio file from the plurality of digital audio files in the audio stream.
Claim 3	
--	A method of rendering digital content according to claim 2, wherein the time information includes at least one of a start time, an end time, and a duration of each digital audio file in the plurality of digital audio files, the start and end times of each digital audio file determined relative to the beginning of the audio recording.
Claim 4	
--	A method of rendering digital content according to claim 3, wherein determining the first digital data file comprises comparing the predetermined time offset to the at least one of a start time, an end time, and a duration of each digital audio file in the plurality of digital audio files to find a first digital audio file that includes digital audio data from a point in the audio stream corresponding to the predetermined time offset.

Listing of Claims from U.S. 9,319,720	
Claim 5	
--	A method of rendering digital content according to claim 4, wherein downloading the first digital data file or the digital data segment contained in the first digital data file and determined using the time information in the descriptor file comprises downloading the first digital audio file from the at least one server so that the first digital audio file is resident with the media player, and wherein rendering the digital content using the media player comprises rendering the first digital audio file using the media player while simultaneously downloading subsequent digital audio files from the at least one server.
Claim 6	
--	A method of rendering digital content according to claim 4, wherein the plurality of media streams further comprises a text stream including a plurality of electronic text files created by segmenting an electronic text of the originating written work, and wherein the time information further includes at least one of a start time, an end time, and a duration of each electronic text file in the plurality of electronic text files, the start and end times of each electronic text file determined relative to the beginning of the audio recording.
Claim 7	
--	A method of rendering digital content according to claim 6, wherein the method further comprises determining a first electronic text file from the plurality of electronic text files in the text stream by comparing the predetermined time offset to the at least one of a start time, an end time, and a duration of each electronic text file in the plurality of electronic text files to find a first electronic text file having a point therein corresponding to the predetermined time offset, and downloading the first electronic text file from the at least one server so that the electronic text file is resident with the media player.

Listing of Claims from U.S. 9,319,720	
Claim 8	
--	A method of rendering digital content according to claim 7, wherein rendering the digital content using the media player comprises rendering the first digital audio file using the media player while simultaneously rendering the first electronic text file using the media player starting from a point in the first electronic text file calculated to correspond to the predetermined time offset.
Claim 9	
--	A method of rendering digital content according to claim 1, wherein the plurality of media streams comprises a text stream including a plurality of electronic text files created by segmenting an electronic text of the originating written work, and wherein the time information includes at least one of a start time, an end time, and a duration of each electronic text file in the plurality of electronic text files, the start and end times of each electronic text file determined relative to the beginning of the audio recording, and wherein the first digital data file is a first electronic text file from the plurality of electronic text files in the text stream.
Claim 10	
--	A method of rendering digital content according to claim 9, wherein determining the first digital data file comprises comparing the predetermined time offset to the at least one of a start time, an end time, and a duration of each electronic text file in the plurality of electronic text files to find a first electronic text file having a point therein corresponding to the predetermined time offset.

Listing of Claims from U.S. 9,319,720	
Claim 11	
--	A method of rendering digital content according to claim 10, wherein downloading the first digital data file or the digital data segment contained in the first digital data file and determined using the time information in the descriptor file comprises downloading the first electronic text file from the at least one server, loading the first electronic text file into a formatted memory structure, and loading a plurality of hard time tags from the descriptor file into the formatted memory structure.
Claim 12	
--	A method of rendering digital content according to claim 11, wherein rendering the digital content using the media player comprises determining at least one of a start time and an end time of a plurality of paragraphs in the loaded first electronic text file using the plurality of hard time tags, the start and end times of each paragraph determined relative to one of the beginning of the audio recording and a beginning of the first electronic text file.
Claim 13	
--	A method of rendering digital content according to claim 11, wherein rendering the digital content using the media player comprises extrapolating time values for a plurality of characters in the loaded first electronic text file, the time values extrapolated from the at least one of a start time, an end time, and a duration of the first electronic text file and the plurality of hard time tags.
Claim 14	
--	A method of rendering digital content according to claim 10, wherein rendering the digital content using the media player comprises determining a current time offset using at least one of a discrete and a timed event.

Listing of Claims from U.S. 9,319,720	
Claim 15	
--	A method of rendering digital content according to claim 14, wherein the plurality of media streams further comprises an audio stream including a plurality of digital audio files created by segmenting the audio recording, and wherein rendering the digital content using the media player further comprises simultaneously rendering a digital audio file from a point in the audio stream corresponding to the predetermined time offset and the first electronic text file from a point in the text stream corresponding to the predetermined time offset.
Claim 16	
--	A method of rendering digital content according to claim 15, wherein rendering the digital content using the media player further comprises providing automated page turns within the first electronic text file using the current time offset.
Claim 17	
--	A method of rendering digital content according to claim 1, wherein the plurality of media streams comprises an ancillary content stream including a plurality of ancillary files, and wherein the time information includes at least one of a start time, an end time, and a duration of each ancillary file in the plurality of ancillary files, the start and end times of each ancillary file determined relative to the beginning of the audio recording, the ancillary content stream including at least one of a plurality of graphic files, a plurality of advertising files, and a plurality of user generated ancillary files.
Claim 18	
--	A method of rendering digital content according to claim 1, wherein the predetermined time offset is provided by one of a bookmark, a search result, a table of contents, an index, a list of tables, a list of figures, footnotes, quotations, a list of illustrations, a fast-forward request, and a rewind request.

Listing of Claims from U.S. 9,319,720	
Claim 19	
--	A method of rendering digital content according to claim 1, wherein the predetermined time offset is provided by a bookmark, the bookmark including information to identify the descriptor file.
Claim 20	
--	A method of rendering digital content according to claim 10, wherein the plurality of media streams further comprises an audio stream including a plurality of digital audio files created by segmenting the audio recording and an illustration stream including a plurality of illustration files from the originating written work, and wherein the descriptor file includes at least one of a start time, an end time, and a duration of each digital audio file in the plurality of digital audio files, and at least one of a start time, an end time, and a duration of each illustration file in the plurality of illustration files, the start and end times of each digital audio file and each illustration file determined relative to the beginning of the audio recording.
Claim 21	
--	A method of rendering digital content according to claim 10, wherein the predetermined time offset is provided from a text search result.
Claim 22	
--	A method of rendering digital content according to claim 14, further comprising creating a bookmark for a selected point in the digital content, the bookmark including the current time offset, the current time offset determined relative to the timeline of the audio recording.
Claim 23	
23[a]	A method of rendering digital content according to claim 1, further comprising: determining a current time offset while rendering the digital content; and

Listing of Claims from U.S. 9,319,720	
23[b]	creating a bookmark for a selected point in the rendered digital content, the bookmark an electronic file including the current time offset and an identification of the descriptor file, the current time offset being a time difference from a beginning of the audio recording to a selected point in the audio recording, the selected point in the audio recording corresponding to the selected point in the digital content.
Claim 24	
24[pre]	A method of rendering digital content comprising:
24[a]	providing a media player having access to at least one server via a network,
24[b]	the at least one server having stored thereon a descriptor file, an audio stream, and a text stream, the audio stream and the text stream derived from a same originating written work,
24[c]	the audio stream including a plurality of digital audio files created by segmenting an audio recording of the originating written work, the text stream including a plurality of electronic text files created by segmenting an electronic text of the originating written work,
24[d]	the descriptor file including information allowing a simultaneous synchronized rendering of the audio stream and the text stream that provides a literary experience of the originating written work,
24[e]	the information including time information for each digital audio file in the audio stream and for each electronic text file in the text stream, the time information for synchronizing the audio stream and the text stream and determined relative to a timeline of the audio recording;

Listing of Claims from U.S. 9,319,720	
24[f]	determining a first digital audio file from the plurality of digital audio files in the audio stream, the first digital audio file having audio content to be rendered with the media player, and determined using the time information in the descriptor file and a predetermined time offset, the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording;
24[g]	downloading the first digital audio file or a digital audio segment contained in the first digital audio file and determined using the time information in the descriptor file from the at least one server so that the audio content is resident with the media player;
24[h]	determining a first electronic text file from the plurality of electronic text files in the text stream, the first electronic text file having text content to be rendered with the media player and determined using the time information in the descriptor file and the predetermined time offset;
24[i]	downloading the first electronic text file or an electronic text segment contained in the first electronic text file and determined using the time information in the descriptor file from the at least one server so that the text content is resident with the media player; and
24[j]	simultaneously rendering the audio content and the text content using the media player at an arbitrary point determined using the predetermined time offset,
24[k]	comprising providing automated page turns for the text content, the automated page turns determined in dependence upon a progressing time offset relative to the rendering of the audio content.
Claim 25	
25[pre]	A system for rendering digital content comprising:
25[a]	at least one server having stored thereon a descriptor file and a plurality of media streams derived from a same originating written work,

Listing of Claims from U.S. 9,319,720	
25[b]	each media stream including a plurality of digital data files,
25[c]	the descriptor file including information allowing a simultaneous synchronized rendering of the plurality of media streams that provides a literary experience of the originating written work,
25[d]	the information including time information for each digital data file in each media stream, the time information for synchronizing the plurality of media streams and determined relative to a timeline of an audio recording of the originating written work;
25[e]	a client including computer readable code for:
25[f]	determining a first digital data file from the plurality of digital data files in a first media stream from the plurality of media streams, the first digital data file having digital content to be rendered with the media player and determined using the time information in the descriptor file and a predetermined time offset, the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording;
25[g]	downloading the first digital data file or a digital data segment contained in the first digital data file and determined using the time information in the descriptor file from the at least one server so that the digital content is resident with the client; and
25[h]	rendering the digital content at an arbitrary point determined using the predetermined time offset; and
25[i]	a network for connecting the at least one server and the client.
Claim 26	
26[pre]	A non-transitory computer readable storage medium including computer readable code, which when executed by a computer, causes said computer to:

Listing of Claims from U.S. 9,319,720	
26[a]	access at least one server via a network,
26[b]	the at least one server having stored thereon a descriptor file and a plurality of media streams derived from a same originating written work,
26[c]	each media stream including a plurality of digital data files, the descriptor file including information allowing a simultaneous synchronized rendering of the plurality of media streams that provides a literary experience of the originating written work,
26[d]	the information including time information for each digital data file in each media stream, the time information for synchronizing the plurality of media streams and determined relative to a timeline of an audio recording of the originating written work;
26[e]	determine a first digital data file from the plurality of digital data files in a first media stream from the plurality of media streams, the first digital data file having digital content to be rendered with the media player and determined using the time information in the descriptor file and a predetermined time offset, the predetermined time offset external to the descriptor file and determined relative to the timeline of the audio recording;
26[f]	download the first digital data file or a digital data segment contained in the first digital data file and determined using the time information in the descriptor file from the at least one server so that the digital content is resident with the computer: and
26[g]	render the digital content at an arbitrary point determined using the predetermined time offset.
Claim 27	
27[pre]	A non-transitory computer readable storage medium according to claim 26, which when executed by a computer, further causes said computer to:

Listing of Claims from U.S. 9,319,720	
27[a]	determine a current time offset while rendering the digital content; and
27[b]	create a bookmark for a selected point in the rendered digital content, the bookmark an electronic file including the current time offset and an identification of the descriptor file, the current time offset being a time difference from a beginning of the audio recording to a selected point in the audio recording, the selected point in the audio recording corresponding to the selected point in the digital content.
Claim 28	
--	A non-transitory computer readable storage medium according to claim 27, which when executed by a computer, further causes said computer to electronically transfer the bookmark to another computer.
Claim 29	
--	A non-transitory computer readable storage medium according to claim 27, which when executed by a computer, further causes said computer to transfer one of the bookmark and a shortcut to the bookmark to one of a desktop, a window other than the window used for rendering the digital content, and an application other than the application used for rendering the digital content.

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. §42.24(d), the undersigned certifies that this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,319,720** contains 13,864 words according to the word-processing program used to prepare this paper. The foregoing word count complies with the 14,000-word type-volume limit specified by 37 C.F.R. §42.24(a)(1).

Dated: March 24, 2025 By: */Colin B. Heideman /*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on the date below a copy of this **PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,319,720** and **ACCOMPANYING EXHIBITS** are being served on March 24, 2025 via Federal Express overnight mail on counsel of record for U.S. Patent No. 9,319,720 at the Correspondence Address of record below:

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

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

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

Dated: March 24, 2025 By: */Colin B. Heideman /*
Colin B. Heideman (Reg. No. 61,513)
KNOBBE MARTENS OLSON & BEAR, LLP